AFOSR-TR-97-0643

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** 23 Jan 97	**3. REPORT TYPE AND DATES COVERED** FINAL TECHNICAL REPORT 11/1/93 - 10/31/96

4. TITLE AND SUBTITLE
ALGORITHMS FOR DIGITAL MICRO-WAVE RECEIVERS & OPTIMAL SYSTEM IDENTIFICATION

5. FUNDING NUMBERS
F49620-94-1-0033

6. AUTHOR(S)
PROFESSOR ARNAB SHAW

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
WRIGHT STATE UNIVERSITY
ELECTRICAL ENGINEERING
DAYTON OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
110 DUNCAN AVENUE SUITE B115
BOLLING AFB DC 20332-8050

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

19971203 171

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE, DISTRIBUTION IS UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

According to the original Project Proposla, the tow primary directions considered in this research effort are;
(i) Advanced signal processing algorithms for digital microwave receivers with Electronic Warfare applications: Significant contribution has been made on estimating the Angles-Of-Arrival (AOA) or frequencies. Specifically, a computationally efficient and accurate Minimum-Norm Method has been developed that does not require any Eigenanalysis. Theoretical Perturbation Analysis of this method has been completed. A Maximum-Likelihood Estimator (MLE) that ensures unit circle frequencies has been developed. Furthermore, a new pipelined adaptive algorithm for Tracking moving targets has been presented.
(ii) Optimal identification of rational transfer functions: unlike existing algorithms which either modify or linearize the erro criterion, the true criteria have been decoupled into (i) a purely linear problem for numerator and (ii) a nonlinear problem with reduced dimensionality for the denominator. the decoupled estimators possess global optimality properties buth have reduced computational complexity than existing methods. The results on 1-D cases have been extended for idtntifying Multi-Dimensional systems.
In addition, a new "Distributed Look-Ahead" architecture and an Optimal Approximation Approach have been proposed for high-speed implementation of Recursive Ditigal Filters.

14. SUBJECT TERMS

15. NUMBER OF PAGES
167

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
U	U	U	

TABLE OF CONTENTS

page

CHAPTER 1 : EXECUTIVE SUMMARY 3

List of Publications and Manuscripts under Review/Preparation 8

Personnel Invloved 10

CHAPTER 2 : THE DIGITAL MICROWAVE RECEIVER DESIGN PROBLEM 11

Section - 2.1 : Superresolution without Eigendecomposition : Method and Perturbation Analysis 13

Section - 2.2 : Maximum Likelihood Estimation of Multiple Frequencies with Constraints to Guarantee Unit Circle roots 39

Section - 2.3 : Improved AR-Parameter Estimation From Noisy Observation Data 47

Section - 2.4 : Improved ARMA-Parameter Estimation From Noisy Observation Data 54

Section - 2.5 : Time-Domain Detection of Electronic Warfare Signals in Noise 62

Section - 2.6 : Pipelined-Adaptive Tracking of Multiple Sinusoidal Frequencies 68

CHAPTER 3 : SYSTEM IDENTIFICATION AND HARDWARE IMPLEMENTATION PROBLEMS .. 76

Section - 3.1 : Identification of 1-D Rational Systems from Input-Output Data 78

Section - 3.2 : Identification of 1-D Rational Systems in the Frequency Domain 86

Section - 3.3 : Optimal Estimation of the Parameters of All-Pole Transfer Functions 100

Section - 3.4 : Design of Denominator Separable 2-D IIR Filters 120

Section - 3.5 : Optimal Frequency Domain Design of Denominator Separable Two-Dimensional Digital IIR Filters 132

Section - 3.6 : Optimal Spatial-Domain Design of 2-D IIR Filters 138

Section - 3.7 : Distributed Look-Ahead : A General Approach for Pipelining Recursive Digital Filters 145

Section - 3.8 : Optimal Least-Squares Design of Pipelined Recursive Filters in the Time-Domain 154

Section - 3.9 : Pipelined Look-Ahead Implementation of a Class of 2-D IIR Filters 161

CHAPTER 1

Executive Summary

Microwave receivers play a vital role in Electronic Warfare (EW) environments for *passive* identification and localization of unknown targets emitting high-frequency electro-magnetic signals. These Receivers process signals received by Microwave band radars and majority of these receivers utilize analog signal processing tools and techniques. Microwave signals have very high frequency content and have wide bandwidths. As of now, there are no EW receivers that process microwave radar signals entirely in the digital domain. It is expected that, with the emergence of increasingly faster and inexpensive digital computers and high-speed A/D converters, digital processing of microwave signals would most certainly be the way of the future. One of the main purpose of this project had been to complement the research on Digital Microwave Receiver Design being conducted at the EW Laboratory at WPAFB, Dayton, Ohio.

In addition to the digital receiver design problem, some fundamental theoretical aspects of several classical System Identification problems as well as high-speed implementation of various Signal Processing algorithms have also been addressed as part of this project. In particular, a unified framework has been developed for optimal estimation of rational transfer function parameters from prescribed Time-Domain or Frequency-Domain specifications. This powerful unifying theoretical framework for System Identification appear to have remained mostly unrecognized and un-utilized. Apart from the digital EW receiver design problem, the proposed theoretical foundation is expected to have a broad range of applications in rational modeling.

High-speed implementation of digital signal processing algorithms on Multiprocessor Architecture is an important topic of current interest in recent Signal Processing literature. In EW applications as well as in other hardware implementations of digital systems, high-speed architecture would no doubt play an important role. In view of this, high-speed hardware implementation of a few important signal processing algorithms have been addressed as the final part of this project.

As noted above, significant progress has been made during the course of this research project. Several problems of current interest have been addressed and solved satisfactorily. Most of the new results were proposed in the original proposal although some intermediate work had been undertaken as the needs arose at the Wright Labs. This Final Technical Report contains the details of all the results of the research that have been accomplished over the entire period covered by the project. It may be noted that some of the results presented in this report were initiated as part of the work on the original proposal (Grant No. AFOSR-F49620-93-1-0014).

The importance of any research may perhaps be best judged by the quality of publications it generates. Consequently, a significant amount of time has been devoted on preparing Journal and Conference articles in order to report the findings of this project. Most of the results contained in this report have been published/accepted/presented in internationally recognized and top quality Signal Processing Journals/Conferences although some recent results are currently under review/preparation for future publication. The papers/publications ensuing from this research are listed at the end of this introductory Chapter. Copies of the papers and publications can be made available to the Program Monitor, if desired.

The research conducted under this project can be categorized primarily into two broad themes, *viz.*,

(i) **Digital EW Receiver Design Problems** : The problems addressed are as follows :

  (a) A high-resolution method for AOA estimation using Minimum-Norm Method that does not rely on any Eigendecomposition

(b) Statistical Perturbation Analysis of the DFT-based Minimum-Norm Method proposed in part-a.

(c) A high-resolution Maximum-Likelihood method for frequency estimation that guarantees unit-circle roots

(d) Two methods for superior estimation of AR and ARMA parameters when the observation data is noisy

(e) Time-Domain algorithms for detection of Electronic Warfare Signals in the presence of Noise

(f) Pipelined-Adaptive Tracking of Multiple Sinusoidal Frequencies.

(ii) **System Identification and Hardware Implementation Problems** :

(a) Optimal identification of 1-D Rational Systems from Input-Output Data

(b) Optimal identification of 1-D Rational Systems in the Frequency Domain

(c) Optimal Identification of All-Pole Rational Systems in Time-Domain

(d) Design of Denominator Separable 2-D IIR Filters from Spatial-Impulse Response Data

(e) Design of Denominator Separable 2-D IIR Filters from 2-D Frequency Response Data

(f) Design of 2-D IIR Filters with non-separable denominator from Spatial-Impulse Response Data

(g) High-speed pipelined implementation of 1-D Recursive Filters based on a new *Distributed Look-Ahead* scheme.

(h) Optimal Estimation of LA filters.

(i) High-speed pipelined implementation of 2-D Recursive Filters based on the *Distributed Look-Ahead* scheme proposed in part-g.

The report is organized as follows: In Chapter 2, the research results on Digital EW receiver design related problems are reported whereas in Chapter 3, the System Identification and Hardware Implementation areas are covered with complete details. Individual Chapters are divided into several Sections by topics. In the following paragraphs the main results obtained in these each of these sections are summarized briefly.

CHAPTER 2. THE DIGITAL MICROWAVE RECEIVER DESIGN PROBLEM

Section - 2.1 : Superresolution without Eigendecomposition : Method and Perturbation Analysis : Many existing high-resolution methods, such as MUSIC or Minimum-Norm Method, rely on special-purpose hardware or software for obtaining the signal and noise subspace eigenvectors of Autocorrelation (AC) matrices. In this project, we have developed a new DFT-based high-resolution frequency estimation algorithm which does not require any eigendecomposition and hence, it is much less computation intensive. It has been demonstrated that the DFT of the AC matrix (DFT-of-AC) essentially performs an equivalent task of separating the signal and noise subspaces. Furthermore, when the signal-subspace part of the DFT-of-AC vectors are used in MNM, almost identical high-resolution AOA estimates are produced. The results have been published as a Journal paper [I10] and has been presented at ICASSP-94 [I19]. It may be noted here that according to one of the anonymous reviewers of the journal paper, this work is a "significant breakthrough in source localization".

In the later part of this Section, we present a detailed theoretical Perturbation Analysis of the estimates produced by the D-MNM algorithm. The theoretical results closely corroborate and confirm the superior performance observed in simulations. The results indicate that the high-resolution performance of D-MNM is uniformly superior than its eigen-based counterpart, especially at low SNR. The performance is also superior than the eigen-based root-MUSIC method at low SNR. Furthermore, D-MNM appears to provide better success rate among all methods at low SNR. Close match between the theoretical and simulated performance verifies the validity of the formulae derived here. Preliminary work has been presented at ASILOMAR-94 [I18] and a detailed version is under preparation for a Journal paper [I25].

Section - 2.2 : Maximum-Likelihood Method with Exact Constraints : A recently proposed approximate Maximum-Likelihood Estimator (MLE) of multiple exponentials, converts the frequency estimation problem into a problem of estimating the coefficients of a z-polynomial with roots at the desired frequencies. Theoretically, the roots of the estimated polynomial should fall on the unit circle. But MLE, as originally proposed, does not guarantee unit circle roots. This drawback sometimes causes merged frequency estimates, especially at low SNR. If all the sufficient conditions for the z-polynomial to have unit circle roots are incorporated, the optimization problem becomes too nonlinear and it loses the desirable weighted-quadratic structure of MLE. In this work, the exact constraints are imposed on each of the 1st-order factors corresponding to individual frequencies for ensuring unit circle roots. The constraints are applied during optimization *alternately* for each frequency. In the absence of any merged frequency estimates, the RMS values more closely approach the theoretical Cramer-Rao (CR) bound at low SNR levels. The work has been published as a Journal paper [I5].

Section - 2.3 : Improved AR-Parameter Estimation From Noisy Observation Data : Auto-Regressive (AR) modeling is the most widely used approach for model-based spectrum estimation. But almost all the existing methods for AR-parameter estimation show severe degradation if the observed signal is corrupted with noise. In fact, all the commonly used techniques, such as, Autocorrelation Method (AM), Covariance Method (CM), Modified Covariance Method and their variations, give poor Power Spectral Density (PSD) estimates when the observations are noisy. In this part of the project, a *data-adaptive pre-filtering* approach is presented to address this problem. The results indicate that when only noisy data is available for modeling, the proposed technique gives more accurate PSD estimates than the commonly used methods. A conference paper on this work have been accepted [I21] and a more comprehensive version is under preparation for publication as a Journal paper.

Section - 2.4 : Improved ARMA-Parameter Estimation From Noisy Observation Data : Existing methods for ARMA modeling assume that the available process is produced by an ARMA system driven by a white input process, *i.e.*, the observed process is considered to be pure ARMA. In practice, the available data usually have observation noise added to it but the ARMA methods do not address this problem. Simulations show that performance of the existing ARMA methods deteriorate when the observation process is noisy. In this part of the project a new ARMA algorithm is given which utilizes a recently developed deterministic rational system identification method (OM-IO) that minimizes the modeling or output error norm. The algorithm first estimates the input process and then invokes OM-IO using the input-output data. Simulations indicate that the proposed method is quite effective even at low SNR observation data. A conference paper on this work has been accepted [I20] and a more comprehensive version is under preparation for publication as a Journal paper.

Section - 2.5 : Time-Domain Detection of Electronic Warfare Signals in Noise : Almost all existing AOA/RF estimation algorithms assume that the signal is already present in the observed data. But in the passive mode of operations of EW applications, source signals may not be present at all within the observation window, or the signals may fill only a part of the estimation window. In either case, any frequency estimation algorithm would essentially produce erroneous or noise frequencies because the observed signal would not satisfy the model assumed by the estimation algorithm. Considering the relatively high computational burden, any estimation method should be invoked only when a detection scheme indicates high probability of presence of threat. In this part of the project, the theory of detecting sinusoids from Quantized and Noisy time-domain observation samples have been developed. The theoretical work on single/multiple samples is mostly complete. Studies with

Quantized data have also been performed and the results appear reasonably good. Lab tests for the Envelope Detection and Square-Law cases have been conducted at Wright Labs with satisfactory results.

Section - 2.6 : Pipelined-Adaptive Tracking of Multiple Sinusoidal Frequencies : New Pipelined-Adaptive algorithms are proposed for tracking multiple Frequencies or Angles-of-Arrival (AOA) of moving targets. Pipelining of adaptive filters pose a critical challenge because of the timing mismatch arising from the feedback signals. In this work, some relaxation techniques have been utilized to pipeline adaptive algorithms for high-speed tracking of frequency/AOAs. Two adaptive tracking algorithms have been mapped into pipelined forms, namely Least-Mean Squares (LMS) and Recursive Least-Squares (RLS). Preliminary results have been presented at ISCAS-96 [I15] and a Journal version is under preparation for possible publication [I29].

CHAPTER 3. SYSTEM IDENTIFICATION AND HARDWARE IMPLEMENTATION PROBLEMS

Fundamental contributions have been made in 1-D and 2-D Rational System Identification theory. Several key journal papers have been published/accepted and a number of conference publications have also been generated. The proposed comprehensive framework encompasses a large class of Identification problems including, (a) Input-Output data [I8, I24], (b) Impulse Response Data : AR case [I9, I23], ARMA case [I11] and (c) Frequency Response Data [I3, I22], (d) Multivariable System Identification [I7] and also for shaping Time responses of Minimum Phase Systems [I4]. Key results are summarized below.

Section - 3.1 : Identification of 1-D Rational Systems from Input-Output Data : A theoretical and algorithmic framework is proposed for optimal identification of rational transfer function parameters of discrete-time linear systems from Input-Output (IO) data. The nonlinear criterion is theoretically *decoupled* into a purely linear problem for estimating the optimal numerator and a nonlinear problem for the optimal denominator. The proposed decoupled approach has reduced computational requirements when compared to existing algorithms which estimate the parameters simultaneously. This research has led to one Journal paper [I8] and a Conference paper [I24].

Section - 3.2 : Identification of 1-D Rational Systems in the Frequency Domain : A new Frequency-Domain (FD) approach has been developed for optimal estimation of rational transfer functions coefficients. The proposed method seeks to match any arbitrarily-shaped FD specifications in the Least-Squares (LS) sense. The desired specifications may be arbitrarily spaced in frequency. The design is performed directly in the digital domain and no analog to digital transformation is necessary. The proposed method makes use of the inherent mathematical structure in this rational modeling problem to theoretically decouple the numerator and denominator estimation problems into two smaller dimensional problems. The denominator criterion is nonlinear but possesses a weighted-quadratic structure which is convenient for iterative optimization. The optimal numerator is found linearly by solving a set of simultaneous equations. The decoupled criteria retain the global optimality properties. The performance of the algorithm is demonstrated with some simulation examples. This research has led to one Journal paper [I3] and a Conference paper [I22].

Section - 3.3 : Identification of All-Pole Rational Systems in Time-Domain : An algorithm is proposed for optimal estimation of the parameters of Auto-Regressive (AR) or all-pole transfer function models from prescribed impulse response data. The transfer function coefficients are estimated by minimizing the ℓ_2-norm of the exact model fitting error. Existing methods either minimize equation errors or modify the true non-linear fitting error criterion. In the proposed method, the multidimensional nonlinear error criterion has been decoupled into a purely linear and a nonlinear subproblem. Global optimality properties of the decoupled estimators have been established. For data corrupted with Gaussianly distributed noise, the proposed method produces

Maximum-Likelihood Estimates (MLE) of the AR-parameters. The inherent mathematical structure in the nonlinear subproblem is exploited in formulating an efficient iterative computational algorithm for its minimization. The proposed algorithm provides an useful computational tool based on appropriate theoretical foundation for accurate modeling of all-pole systems from prescribed impulse response data. The effectiveness of the algorithm has been demonstrated with several simulation examples. This research has led to one Journal paper [I9] and a Conference paper [I23].

Section - 3.4 : Design of Denominator Separable 2-D IIR Filters : This work extends the 1-D results in [I11] to 2-D system identification. In this part of the report, the optimal design of an important class of two-dimensional (2-D) digital IIR filters from spatial impulse response data is addressed. The denominator of the desired 2-D filter is assumed to be separable into two 1-D factors. The filter coefficients are estimated by minimizing the ℓ_2-norm of the error between the prescribed and the estimated spatial domain responses. The denominator and numerator estimation problems are theoretically decoupled into separate problems. The decoupled criteria have reduced dimensionality. The denominator criterion is simultaneously optimized *w.r.t.* the coefficients in both dimensions using an iterative algorithm. The numerator coefficients are found in a straightforward manner. If the desired response is known to be symmetric, the proposed algorithm can be constrained to have separable-denominators. Initial results have been published as a Journal paper [I6] and some further developments are currently being considered [I2].

Section - 3.5 : Optimal Frequency Domain Design of Denominator Separable Two-Dimensional Digital IIR Filters : Classical design techniques using Butterworth, Chebyshev or Elliptic polynomial are only limited particular types of design specifications, such as Bandpass, lowpass etc. A least-squares technique is presented for designing quarter-plane separable-denominator 2-D IIR filters to best approximate prescribed frequency domain (FD) specification of any arbitrary shape. Structured Matrix Approximation approach is utilized to show that the FD error vector is linearly related to the 2-D numerator coefficients whereas the relationship with the 2-D denominators is quasi-linear. Furthermore, the numerator and denominator estimation problems are theoretically decoupled. The quasi-linear relationship with the denominator is used to formulate an algorithm for iterative estimation of the denominator. The numerator is found in one step using the estimated denominator. Computer simulations show the effectiveness of the proposed method and its superior performance compared to several existing methods. This work has been presented at ICASSP-95 [I17]. A detailed version is also under preparation for a Journal paper [I30].

Section - 3.6 : Optimal Spatial-Domain Design of 2-D IIR Filters : In this Section we present a structured matrix approximation framework to develop the most general form for *optimal* least-squares (LS) design of 2-D recursive filters from prescribed spatial domain data. Unlike the work in Section 3.4, no separability is assumed for the 2D denominator. Utilizing matrix structures inherent in this problem it is shown that the exact ℓ_2 error has a purely *linear* relationship with the 2-D numerator parameters whereas the 2-D denominator coefficients are nonlinearly related to the error. But more interestingly, the denominator and numerator estimation problems are theoretically decoupled into separate problems without affecting any optimality properties. In the decoupled form, the numerator estimation problem is shown to be purely linear. For estimating the denominator also, it is shown that the decoupled ℓ_2 error vector possesses a *quasi-linear* relationship with the denominator coefficients. Decoupled estimation leads to reduced computational complexity because there is no need for iterating on the numerators. Simulation results indicate that for several common filer design problems, the proposed general version performs better than the separable design developed earlier in Section 3.4. Preliminary results have been presented at ISCAS-95 [I15] and a Journal version is under consideration for possible publication [I1].

Section - 3.7 : Distributed Look-Ahead : A General Approach for Pipelining Recursive Digital Fil-

ters : A new Look-Ahead (LA) scheme, *Distributed Look-Ahead* (DLA), is proposed for pipelined implementation of recursive digital filters. It is established that in case of many recursive filters, DLA can provide equivalent and *stable* implementation with reduced pipeline delay and hardware complexity, when compared with some existing LA schemes. The existing Scattered Look-ahead implementation achieves stability at the cost of increased multiplication and latch complexities and considerable delay in output generation. The Clustered look-ahead approach can not always guarantee stability. This work shows that, in order to attain stability, the output samples need not be clustered or equally scattered. Indeed, in many filter design problems, stability can be maintained by using *unequally distributed* past output samples. When compared with the scattered approach, the proposed scheme uses fewer number of pole-zero cancelations and the introduced roots are not necessarily at the same radii as the original filter poles. Hence, the proposed DLA scheme has reduced multiplication and latch complexities, higher area-efficiency and it produces outputs with reduced delays. Preliminary results have been presented at ICASSP-96 [I13] and ISCAS-96 [I14] and a Journal version is under preparation for possible publication [I26].

Section - 3.8 : Optimal Least-Squares Design of Pipelined Recursive Filters in the Time-Domain : Currently, look-ahead (LA) pipelined recursive filters are obtained primarily via transformation of a *given* un-pipelined transfer function. For these approaches, it is assumed that the un-pipelined transfer function has already been designed as an intermediate step. In this Section, we present a new algorithm (OM-LA) for *direct* and *optimal* estimation of the coefficients of recursive filters in look-ahead pipelined form. OM-LA is developed by appropriate modification of a recently proposed optimal method (OM) for designing un-pipelined filters (developed previously by the PI as part of a project supported by the AFOSR). It is demonstrated that the proposed one-step approximation can achieve superior match with reduced pipelined filter order because it does not rely on pole-zero cancelations as in current LA pipelining approaches. It is also shown that the denominator polynomial can be constrained to possess any of the possible look-ahead configurations. Unlike some existing methods, OM-LA minimizes the *true* time-domain fitting error-norm between the prescribed and the estimated impulse response and produces superior results. Preliminary results have been presented at ICASSP-96 [I12] and a Journal version is under preparation for possible publication [I27].

Section - 3.9 : Pipelined Look-Ahead Implementation of a Class of 2-D IIR Filters : In Section-3.7, we have presented a new scheme (referred to as *distributed look-ahead*) which is a compromise between the two existing look-ahead approaches for high speed implementation of 1-D Recursive Digital filters. To date neither the Scattered Look-ahead nor the Distributed scheme has so far been utilized for 2-D IIR filter implementation, primarily because the 1-D stability properties of these LA schemes do not necessarily translate to general 2-D IIR filters. The primary focus of this paper is to demonstrate that for a special but very important class of 2-D IIR filters, namely for Denominator Separable configurations, the benefits of these stable look-ahead schemes can indeed be taken advantage of. The results will be submitted for review to ASILOMAR-97 which will be held in November at Naval Postgraduate School [I28]. A detailed version is also under preparation for a Journal paper [I31].

Journal and Conference Articles - Published/accepted/under review

[I1] A. K. Shaw and S. Pokala, "A Structured Matrix Approach for Spatial Domain Design of 2-D IIR Filters," *IEEE Transactions on Circuits and Systems*, accepted for publication Aug., 1996.

[I2] A. K. Shaw and S. Pokala, "Spatial Domain Design of Denominator Separable Multidimensional IIR Filters," *Multidimensional Systems and Signal Processing*, under 2nd review, Sep., 1995.

[I3] A. K. Shaw, "Optimal Design of Digital IIR Filters by Model-Fitting Frequency Response Data," *IEEE*

Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 42, no. 11, pp. 702-710, Nov., 1995.

[I4] P. Misra and A. K. Shaw, "Shaping Time Response by State Feedback in Minimum-Phase Systems," *Journal of Control, Guidance, and Dynamics*, vol. 18, no. 4, pp. 913-916, Jul.-Aug., 1995.

[I5] A. K. Shaw, "Maximum Likelihood Estimation of Multiple Frequencies with Constraints to Guarantee Unit Circle Roots," *IEEE Transactions on Signal Processing,* vol. 43, no. 3, pp. 796-799, Mar., 1995.

[I6] A. K. Shaw, "Design of Denominator Separable 2-D IIR Filters," *Signal Processing,* Switzerland, vol. 42, no. 1, pp. 191-206, Feb., 1995.

[I7] A. K. Shaw, P. Misra and R. Kumaresan, "Multi-Dimensional System Identification From Impulse Response Data," *Circuits, Systems and Signal Processing - Special Issue on Multivariable Systems*, vol. 13, no. 6, pp. 759-782, Dec., 1994.

[I8] A. K. Shaw, "A Decoupled Approach for Optimal Estimation of Transfer Function Parameters from Input-Output Data," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1275-1278, May, 1994.

[I9] A. K. Shaw, "Optimal Estimation of the Parameters of All-Pole Transfer Functions," *IEEE Transactions on Circuits and Systems*, vol. 41, no. 2, pp. 140-150, Feb., 1994.

[I10] A. K. Shaw and W. Xia, "Minimum-Norm Method Without Eigendecomposition," *IEEE Signal Processing Letters*, vol. 1, no. 1, pp. 12-14, Jan., 1994.

[I11] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 113-120, Jan., 1994.

[I12] S. Pokala, A. K. Shaw and M. Imtiaz, "Optimal Least-Squares Design of Pipelined Recursive Filters in the Time-Domain," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, Georgia, May, 1996.

[I13] A. K. Shaw and M. Imtiaz "New Look-Ahead Algorithm for Pipelined Implementation of Recursive Digital Filters," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, Georgia, May, 1996.

[I14] A. K. Shaw and M. Imtiaz "A General Look-Ahead Algorithm for Pipelining IIR Filters," *International Symposium on Circuits and Systems*, Atlanta, Georgia, May, 1996.

[I15] M. Imtiaz and A. K. Shaw "Tracking of Multiple Targets using Pipelined-Adaptive Algorithm," *International Symposium on Circuits and Systems*, Atlanta, Georgia, May, 1996.

[I16] S. Pokala and A. K. Shaw, "Optimal Spatial Domain Design of 2-D IIR Filters," *IEEE International Symposium on Circuits and Systems*, Seattle, Washington, pp. I335-I338, April, 1995.

[I17] S. Pokala and A. K. Shaw, "Optimal Frequency Domain Design of Denominator Separable Two-Dimensional Digital IIR Filters," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, Michigan, pp. 2133-2136, May, 1995.

[I18] A. K. Shaw and W. Xia, "DFT-Based Preprocessing for High-Resolution Angles-of-Arrival Estimation Without Eigendecomposition," *Twenty-Seventh ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 826-830, Oct., 1994.

[I19] A. K. Shaw and W. Xia, "High-Resolution Angles of Arrival Estimation using Minimum-Norm Method Without Eigendecomposition," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Adelaide, Australia, vol. IV, pp. 233-236, April, 1994.

[I20] A. K. Shaw and S. Kundu, "Improved ARMA Modeling from Noisy Observations," *Twenty-Seventh ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct., 1993.

[I21] A. K. Shaw and S. Kundu, "AR-Spectrum Estimation from Noisy Observation Data," *Twenty-Seventh ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct., 1993.

[I22] A. K. Shaw, "Optimal Design of Digital IIR Filters by Model-Fitting Frequency Response Data," *IEEE International Symposium on Circuits and Systems*, Chicago, IL, pp. 475-478, May, 1993.

[I23] A. K. Shaw, "Optimal Estimation of AR-Model Parameters from Impulse Response Data," *31st IEEE Conference on Decision and Control*, Tucson, AZ, pp. 903-908, Dec., 1992.

[I24] A. K. Shaw, "A New Algorithm for Optimal Estimation of Plant Parameters from Input-Output Data," *31st IEEE Conference on Decision and Control*, Tucson, AZ, pp. 1684-1685, Dec., 1992.

[I25] A. K. Shaw and W. Xia, "Superresolution without Eigendecomposition : Method and Perturbation Analysis," Under Preparation for *IEEE Transactions on Aerospace and Electronic Systems.*

[I26] A. K. Shaw and M. Imtiaz, "Distributed Look-Ahead : A General Approach for Pipelining Recursive Digital Filters," Under Preparation for *IEEE Transactions on Circuits and Systems.*

[I27] A. K. Shaw and M. Imtiaz, "Optimal Least-Squares Design of Pipelined Recursive Filters in the Time-Domain," Under Preparation for *IEEE Transactions on Circuits and Systems.*

[I28] A. K. Shaw and M. Imtiaz, "Look-Ahead Pipelined Implementation of a Class of 2-D IIR Filters," Under Preparation for *Thirty-First ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct., 1997.

[I29] A. K. Shaw and M. Imtiaz, "Pipelined-Adaptive Tracking of Multiple Sinusoidal Frequencies," Under Preparation for *IEEE Signal Processing Letters.*

[I30] S. Pokala and A. K. Shaw, "Frequency Domain Design of Two-Dimensional Digital IIR Filters : Denominator Separable Case," Under Preparation for *IEEE Transactions on Circuits and Systems.*

[I31] A. K. Shaw and M. Imtiaz, "Pipelined Look-Ahead Implementation of Denominator-Separable 2-D IIR Filters," Under Preparation for *IEEE Transactions on Circuits and Systems.*

Personnel Invloved

Dr. Arnab K. Shaw (PI)

Mr. Wei Xia (Graduate Student)

Mr. Srikanth Pokala (Graduate Student)

CHAPTER 2

THE DIGITAL MICROWAVE RECEIVER DESIGN PROBLEMS

Introduction

Digital processing of microwave signals in Electronic Warfare (EW) environment poses a great challenge to researchers in Signal Processing. Along with the standard requirements of any conventional radar, EW receiver design problem is complicated by the fact that no knowledge about the input signal is available to the receiver. The nature of the problem also requires that measurements and decisions be taken immediately or within a few seconds in an entirely passive mode of operation. All microwave receivers used in practice utilize analog signal processing techniques. The frequency-band of the EW signals are in the GHz range and the signals have wide bandwidths which necessitate sampling and processing of a massive amount of data at or near real-time. Presently, no EW receiver processes microwave radar signals entirely in the digital domain. But it is expected that with the emergence of increasingly faster and inexpensive digital computers and high-speed A/D converters, digital processing of microwave signals would most certainly be the way of the future.

In the past two decades, many classes of radar and sonar receivers have been converted from conventional analog technology to purely digital or hybrid systems, but EW receivers are yet to make such a transition. The primary technological factors that have been holding back possible fabrication of any digital EW receiver are probably twofold. Firstly, if Analog-to-Digital (A/D) converters are to be used at the operating frequency range, then the Nyquist rate would necessitate sampling at the GHZ range and secondly, the digital hardware or firmware must have the capacity to process such high data rate and produce effective results at near real-time.

Digital EW receivers can be expected to offer some major advantages over their analog counterparts. Foremost among these is the almost lossless storage capability of digital memories which can eliminate the dependence on lossy analog delay lines. Digital processors and memory chips are relatively inexpensive, compact in size and have low weight and the trends are towards even further reductions. Digital signal processing algorithms and digital computing technology have matured tremendously and offer a wide range of capabilities. Parallel processing, pipelining, RISC, VLSI design, systolic architecture, vectorization and array processing, fault tolerant computing and etc., are only some of the well-known aspects of digital computing that the last few decades of research have produced. As our research progresses, we intend to study if some of these ideas can be incorporated in the digital receiver in order to improve the efficiency and accuracy of its performance.

The primary task of a microwave receiver is to gather data for sorting of signals and for identifying the radar-type. Based on these information, jamming, weapon delivery or other options are considered. In order to perform these tasks, the receiver must analyze the received radar pulses and measure or estimate the following six parameters : Angle-of-Arrival (AOA), Radio Frequency (RF), Time of Arrival (TOA), Pulse Amplitude (PA), Pulse Width (PW) and Polarization (P).

A critical requirement of an EW receiver is the AOA measurement which is known to be a rather difficult multidimensional nonlinear optimization problem, especially when multiple closely-spaced threats are to be resolved. It is also desirable to have high sensitivity and large dynamic range such that a broad range of signals, including weak ones, can be detected.

As part of this project several AOA/frequency estimation algorithms has been developed and studied. Most existing high-resolution frequency-estimation algorithms rely on special-purpose hardware or software, such as, Eigendecomposition or SVD. In Section 2.1, a DFT-based Minimum-Norm method is proposed which does not require any eigendecomposition but produces high-resolution frequency estimates. This new algorithm needs only to compute the DFT of the Autocorrelation matrix to separate the signal and noise subspaces. Hence the computational burden is much lower than existing high-resolution methods. Therefore, this algorithm appears to be very well-suited for EW applications. The Statistical Performance Analysis of this new algorithm has also been performed and the results are included in Section 2.1 also.

Another new class of algorithms, referred to as KiSS/IQML, have been developed recently for obtaining the Maximum Likelihood Estimates (MLE) of frequencies or AOAs from the roots of $z-$polynomials. But the estimated roots are not guaranteed to fall on the unit circle, as desired. Based on the theory on zeros of polynomials, a new scheme is proposed in Section 2.2 here that will ensure unit circle roots. Many frequency estimation methods make use of the property that a sinusoidal process can be modeled as a limiting case of a narrow-band auto-regressive (AR) process. But the performance of all existing AR parameter estimation methods degrade significantly when the observation data is corrupted with noise. A pre-filtering approach is presented in Section 2.3 that can improve AR-parameter estimates from noisy observation data. Another data-adaptive approach for improved modeling of ARMA processes from noisy observations is given in Section 2.4.

Parameter estimation schemes either follow or work in parallel with a detection scheme ensuring the presence of any threat. A combined detection-estimation scheme has the potential to cut-down computational burden on the signal processor. As a part of this project, statistical theory on hypothesis testing has been utilized for detecting whether a threat is present or not. In Section 2.5, the time-domain detection problem has been presented for single and multiple samples. Specifically, the detection thresholds and Probability of Detection based on Neyman-Pearson Criterion have been derived.

The AOA estimation algorithms presented in Sections 2.1-2.2 work on batch mode where the targets are assumed to be "locally stationary". However, in many practical situations the targets may be non-stationary an it is desirable to track its movements adaptively. In these regards, Pipelined-Adaptive algorithms are proposed in Section 2.6 for tracking multiple Frequencies or Angles-of-Arrival (AOA) of moving targets. Pipelining of adaptive filters pose a critical challenge because of the timing mismatch arising from the feedback signals. In this work, some relaxation techniques have been utilized to pipeline adaptive algorithms for high-speed tracking of frequency/AOAs.

Section - 2.1 : Superresolution without Eigendecomposition : Method and Perturbation Analysis

Summary

Many existing methods for estimating closely spaced sinusoidal frequencies utilize the eigenvectors of Autocorrelation (AC) matrices [1, 10, 15, 26, 30]. Instead, this work considers the use of the DFT of AC matrices (DFT-of-AC) for extracting the signal and noise subspaces. When the signal-subspace part among the DFT-of-AC vectors are used in the Minimum-Norm method (MNM) framework, almost identical high-resolution frequency estimates are produced. Theoretical Perturbation Analysis of the proposed DFT-based MNM (D-MNM) has also been carried out. The analysis confirms that the estimates are theoretically unbiased and have lower theoretical Mean-Squared Error indicating improved high-resolution performance, especially at low SNR. The primary advantages of extracting signal/noise subspace information from DFT-of-AC are reduced computational and hardware complexity than existing methods that need to perform the Eigendecomposition iteratively.

I : Introduction

In many important practical applications, such as radar, sonar and astronomy etc., the resolution capability of FFT is inadequate. Overcoming the resolution limitation of DFT has been a vigorously researched topic in Signal Processing in the past three decades. The modern methods attain the desired 'High-Resolution' or 'Superresolution' at the cost of considerable computational burden. The existing well-known methods often utilize Eigen-Decomposition (ED), Singular Value Decomposition (SVD) or Maximum Likelihood (ML) method which is based on nonlinear optimization [1-5, 8-10, 14-20, 22-24, 26, 29, 30, 32-37, 42-45, 48, 49]. These algorithms, though highly effective, can only be implemented iteratively because of their inherent nonlinearity, which limits their real-time capabilities.

The primary objective of this paper is to effectively combine the computational simplicity of DFT with the underlying mathematical philosophy of certain high-resolution methods. The desired goal is to achieve high-resolution without any iterative optimization. Specifically, many existing high-resolution techniques, such as the Minimum-Norm method (MNM), extract the signal and noise subspace information from the eigenvectors of the Autocorrelation (AC) matrices [15, 26]. It is shown in this paper that the DFT of the AC-matrix (DFT-of-AC) essentially performs an equivalent task of extracting and decoupling the signal and noise subspace information. Hence, it is proposed that the signal eigenvectors be replaced by the largest-norm DFT-of-AC vectors. It is demonstrated that when the DFT-of-AC vectors with larger norms are used in the MNM framework, mostly better or almost equivalent high-resolution DOA estimates are produced. The bias, mean-squared error and the root locations of the proposed DFT-based-MNM (D-MNM) also compare well with the Eigendecomposition-based MNM (E-MNM). The simulations further show that the high-resolution performance of the D-MNM is more robust at low SNR.

In order to establish theoretical justification of the performance of the D-MNM algorithm, we have also conducted theoretical Perturbation Analysis of the estimates produced by the algorithm. The theoretical study corroborates closely with the superior performance already observed in simulations. The details of the following are included after introducing the method. Firstly, the Bias and the Mean-Squared-Error (MSE) in the estimates of the AOAs are shown to be linearly related to the Bias and MSE in the roots of the D-MNM polynomial which, in turn, are shown to depend on the Bias and MSE for the D-MNM coefficient vector. Then the statistics of the coefficient error are related to those of the observed data and the AC matrix estimate. Finally, all the intermediate results are combined and utilized to find the direct statistical relationship between the AOA errors and the observations. The theoretical results indicate that the high-resolution performance of D-MNM is uniformly superior than its eigen-based counterpart, especially at low SNR. The performance is also superior than the eigen-based root-MUSIC method at low SNR. Furthermore, D-MNM appears to provide better success rate among all methods at low SNR. The theoretical analysis closely follows the performance with simulated data, which verifies the validity of the formulae derived here theoretically.

The major significance of the proposed algorithm is that, no complicated iterative optimization is needed and the signal-subspace information is extracted only by a *single matrix multiplication.* Hence, hardware implementation of D-MNM for real-time high-resolution AOA/Frequency estimation may be feasible with currently available technology. It may be noted here that results on some preliminary simulations on the DFT-based method was presented in [38, 39], though no performance analysis was available at the time.

The paper is organized as follows. In Section II, the AOA estimation problem is defined and in Section III some existing approaches are discussed. Some useful properties of the AC matrix are given in Section IV. Then in Section V, the proposed D-MNM algorithm is described. In Section VI, the details of the Perturbation analysis of D-MNM are presented. Some simulation results are given in Section VII and, finally the paper is wrapped with some concluding remarks in Section VIII.

II : Problem Definition

This paper addresses the problem of estimating of the Directions of Arrival (DOA) of densely spaced narrowband targets. Suppose that p plane waves originating from far-field point sources at distinct directions impinge on a linear array of N equally spaced sensors. The signal sampled simultaneously at m^{th} instant of time at N equally spaced sensors form a 'snapshot' vector defined as,

$$\mathbf{x}_m \triangleq [x_m(0)\ x_m(1)\ \ldots\ x_m(N-1)]^t. \qquad (II.1)$$

In the presence of noise, the observation samples can be written as,

$$x_m(n) = \tilde{x}_m(n) + z_m(n) \qquad (II.2)$$

where, $z_m(n)$ represents the additive observation noise and/or the modeling error and $\tilde{x}_m(n)$ denotes the signal part of the observation, which is given by

$$\tilde{x}_m(n) = \sum_{i=1}^{p} A_m(i) e^{j\frac{2\pi d}{\lambda}(n - \frac{N+1}{2})\sin\theta_i + j\phi_m(i)}, \qquad n = 0, 1, \ldots, N-1 \qquad (II.3)$$

where,

p	:	Number of narrowband sources present
d	:	Spacing between sensor elements
λ	:	Wavelength of radiation of the received signals
θ_i	:	Direction-of-Arrival (DOA) of the i^{th} source
$A_m(i)$	:	Amplitude of the i^{th} source at the m^{th} snapshot
$\phi_m(i)$	:	Phase angle of the i^{th} source at the m^{th} snapshot, Uniformly distributed between $-\pi$ and π.

The noise $z_m(n)$ is assumed to be zero-mean and uncorrelated with the source signals and it has a variance of σ_z^2. The signal model can be written in a more succinct form as,

$$\tilde{x}_m(n) = \sum_{i=1}^{p} A_{im} e^{j\omega_i n} \qquad (II.4)$$

where, ω_i and A_{im} are defined as

$$\omega_i \triangleq \frac{2\pi d}{\lambda} \sin\theta_i \qquad \text{and} \qquad (II.5)$$

$$A_{im} \triangleq A_m(i) e^{-j\frac{2\pi d}{\lambda}\left(\frac{N+1}{2}\right)\sin\theta_i \; + \; j\phi_m(i)}. \qquad (II.6)$$

Further details about the above model may be found in [6]. With the above formulation the model for the observation matrix can be written as,

$$\tilde{\mathbf{X}} \triangleq \mathbf{TA} \qquad (II.7)$$

where,

$$\mathbf{T} \triangleq \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_p} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\omega_1(N-1)} & e^{j\omega_2(N-1)} & \dots & e^{j\omega_p(N-1)} \end{bmatrix}, \qquad (II.8)$$

$$\triangleq [\mathbf{t}_1 \; \mathbf{t}_2 \; \cdots \; \mathbf{t}_p], \qquad (II.9)$$

$$\mathbf{A} \triangleq [\mathbf{a}_1 \; \mathbf{a}_2 \; \dots \; \mathbf{a}_M] \qquad \text{and} \qquad (II.10)$$

$$\mathbf{a}_m \triangleq \begin{bmatrix} A_{1m} \\ A_{2m} \\ \vdots \\ A_{pm} \end{bmatrix} \qquad \text{for } m = 1, 2, \dots, M. \qquad (II.11)$$

For half wavelength spacing between two successive sensors of the line array, $\omega_i = \pi \sin\theta_i$. With M snapshot vectors defined in (II.2), the $N \times M$ observation matrix $\mathbf{X}$ is formed as,

$$\mathbf{X} \triangleq [\mathbf{x}_1 \; \mathbf{x}_2 \; \dots \; \mathbf{x}_M]. \qquad (II.12)$$

Using the observation matrix, the spatial covariance matrix can be estimated as,

$$\hat{\mathbf{C}} \triangleq \frac{1}{M}(\mathbf{X}\mathbf{X}^H) \qquad (II.13a)$$

$$\triangleq \frac{1}{M}\sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H. \qquad (II.13b)$$

The description of the observation and the model is now complete. Given the noisy observation matrix $\mathbf{X}$, the problem under consideration in this paper is to estimate the ω_i's and A_{im}'s. Note that the complex amplitudes can be estimated linearly once the ω_i's are known but the estimation of ω_is poses the greatest difficulty because it is a highly nonlinear optimization problem.

III : Existing Methods

It is apparent from the problem statement that the DOA (θ_i) estimation problem is mathematically equivalent to the Frequency Estimation (ω_i) problem which has been a major research topic in many areas of science. Indeed, in the last couple of hundred years, the search for 'hidden periodicities' from observed data has appeared in varied forms in several seemingly differing disciplines of science.

III.a : The Periodogram and its Resolution Limitation

Ever since its discovery in 1965 [6], the FFT has been the primary tool for estimating Directions of Arrival (DOA) or frequencies of far-field sources from noisy observation data. The software or hardware implementation of FFT is remarkably straight-forward. To date, the periodogram continues to be the most frequently used method for frequency/DOA estimation [21, 27]. In fact, it is well known that for localizing a single target, if the noise in the observed data is Gaussianly distributed, the periodogram [27] produces the maximum likelihood estimate. But in case of multiple targets, the periodogram cannot resolve two frequencies which are separated by less than the bin-width of the FFT. In fact, when the sources are spaced at less than the DFT bin-width, the periodogram fails to distinguish two closely spaced frequencies and only provides a single frequency estimate instead of two. The last statement truly portrays the problem one faces while resolving two closely spaced sinusoids when a relatively short data record is available. Clearly, if any amount of data is available for processing, the periodogram of sufficiently Zero-padded (and possibly Windowed) data will provide reasonably good estimates. But in many problems of practical interest only short data record is available and one has to overcome the periodogram's resolution limitation by resorting to what are commonly known in the signal processing literature as 'High-Resolution' or 'Superresolution' techniques. The major contributions in the higher resolution approaches are highlighted next.

III.b : High-Resolution Methods

A multitude of DOA/Frequency Estimation algorithms, their variations and analysis are available in the literature [1-5, 7-20, 22-26, 28-30, 32-49]. In the following paragraphs only some of the major developments are briefly discussed.

Minimum Variance Method : This was perhaps the earliest high resolution methods which was specifically developed for frequency-wavenumber estimation. In order to improve upon Periodogram's resolution limit, Capon had proposed a Minimum Variance method which is a linear estimator that minimizes the interference at frequencies outside the band of interest [4]. Its performance has been shown to be better than the periodogram estimator but worse than the modeling based estimators [20].

Model-Based Methods : A major motivation for many modern high-resolution frequency estimation methods has come from the desire to achieve more exact models for the sinusoids-in-noise data. In the Parameter Estimation area of statistical time-series analysis, it had been well established that Auto-Regressive (AR) modeling is very appropriate for modeling data with peaky spectra. But in the frequency estimation field also, it had been a common knowledge that data composed of sinusoids in noise tend to have peaky spectra. Consequently, frequency estimation based on AR-modeling has received considerable attention [3, 9, 24, 29].

Depending on how the autocorrelation values are estimated, there are three types of AR parameter estimation methods, namely, Autocorrelation method, Covariance method and Modified Covariance method (also known as the Forward-Backward method). The later two cases are more appropriate for sinusoidal processes because of their implicit relationship with Prony's method which provides perfect frequency estimates when no noise is present. Incidentally, the Maximum Entropy method proposed by Burg [7] and the Linear Prediction based spectral estimator [9] both produce essentially identical frequency estimates as the Covariance method.

When p sinusoids are present and a p^{th} order AR model is used, the frequency estimates are found to be poor at low SNR ($\leq 30dB$). To circumvent this hurdle, larger order ($L > p$) AR model has been proposed [25, 60]. The larger model order tends to accommodate a major part of the interfering noise and thereby reduces the effect of noise in the estimates. The larger-order approach performs poorly below 20dB SNR.

Eigen-Analysis of the Auto-Correlation Matrix of Sinusoid-in-Noise Data : AR modeling based approaches offer better resolution performance than their predecessors but these as well as the earlier methods are basically general spectrum estimation methodologies applied to this specific narrow-band problem. Since the mid-to-late seventies, a whole new class of algorithms are being developed by effective exploitation of the special properties of the autocorrelation matrix of the sinusoids-in-noise data. For $N = p+1$, the eigendecomposition of $\mathbf{C}$ was first utilized by Pisarenko [23] who showed that the z-polynomial formed with elements of the eigenvector corresponding to the smallest eigenvalue has roots at the signal frequencies. Though the idea is elegant, Pisarenko's method performs quite poorly for noisy signals. Pisarenko's approach was later improved upon by Kumaresan [15] where, for $N > p$ cases, all the noise eigenvectors had been utilized. As an alternate approach, it was shown in [15] that the signal subspace eigenvectors can also be utilized to form a noise subspace vector which should have zeros at the signal frequency locations. This was achieved in [15, 26] by formulating a Minimum-Norm criterion which is the framework that will be used in the proposed work.

Another major improvement on Pisarenko's approach was presented by Schmidt [30] and Bienvenue and Kopp [1]. They proposed to combine the eigenvectors corresponding to the $(L-p)$ smaller eigenvalues of $\mathbf{C}$ and used an orthogonality criterion to obtain the frequency estimates. In the literature, this approach is known as the 'MUSIC' method.

It may be pertinent to emphasize here that the approach proposed in this work for extracting signal or noise subspace 'without eigendecomposition' may be combined with either the MNM or the MUSIC framework. The MNM framework has been preferred in the development in Section V because in case of the Minimum-Norm method, the frequencies are found directly from the polynomial roots. On the other hand, a search procedure is necessary in case of MUSIC for estimating the frequencies. The polynomial version of MUSIC, known as 'root-MUSIC', could also be used but in that case the order of the z-polynomial would be twice that of MNM.

Maximum-Likelihood Method : This class of algorithms maximize the likelihood function for the observed data, leading to optimization of a non-linear criterion which can only be performed iteratively. Several different approaches are available in the literature [16-18, 27-29, 33-35, 37, 48] among which the recently proposed Constrained MLE approach developed [37] by the first author appears to offer the most accurate results.

Other Methods and the Motivation for the Proposed Method : As listed in the references, there are a large number other methods that address the high-resolution Frequency/DOA estimation problems. In order to achieve the desired high-resolution capability, all these algorithms utilize some form of eigen-analysis or non-linear optimization, both of which are computationally intensive for real-time applications. The primary objective of this paper is to study whether the computational simplicity of DFT can be effectively combined with the underlying mathematical framework of some of the existing high-resolution methods. The final goal is to achieve high-resolution without any iterative optimization such that real-time implementation may be feasible with existing hardware. The proposed method makes use of the special properties of correlation matrices which are outlined next.

IV : Some Properties of the Autocorrelation Matrix

Since the data described by $(II.3)$ is uncorrelated, zero mean WSS process, the $N \times N$ $(N \geq p)$ covariance matrix $\mathbf{C}$ will have the following matrix decomposition when there is no observation noise,

$$\mathbf{C} = \mathbf{T\Sigma T}^H \qquad (IV.1)$$

where, $\mathbf{\Sigma} \triangleq diag\ (\sigma_1^2\ \sigma_2^2\ \ldots\ \sigma_p^2)$ and σ_i^2 denotes the power of the i-th signal. Note that this ideal $\mathbf{C}$ has rank

p. In this case, the eigen-decomposition of $\mathbf{C}$ can be written as,

$$\mathbf{CV} = [\lambda_1\mathbf{v}_1 \ \cdots \ \lambda_p\mathbf{v}_p \ \ 0 \ \cdots \ 0] = \mathbf{\Lambda V} \tag{IV.2a}$$

$$\triangleq \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p & 0 & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} | & | & \cdots & | & | & \cdots & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p & \mathbf{v}_{p+1} & \cdots & \mathbf{v}_N \\ | & | & \cdots & | & | & \cdots & | \end{bmatrix}. \tag{IV.2b}$$

For observations with noise as defined in $(II.3)$,

$$\mathbf{C} = \mathbf{T\Sigma T}^H + \sigma_z^2\mathbf{I}. \tag{IV.3}$$

Note that this theoretical $\mathbf{C}$ has rank N though the signal part, $\mathbf{T\Sigma T}^H$ has rank p. In this case, the eigen-decomposition of $\mathbf{C}$ can be written as,

$$\mathbf{CV} = [(\lambda_1 + \sigma_z^2)\mathbf{v}_1 \ \cdots \ (\lambda_p + \sigma_z^2)\mathbf{v}_p \ \ \sigma_z^2\mathbf{v}_{p+1} \ \cdots \ \sigma_z^2\mathbf{v}_N] \tag{IV.4}$$

where, the λ_i's and σ_z^2 represent the signal and noise eigenvalues. But in practice, the eigendecomposition has to be performed on the sample covariance matrix $\mathbf{C}$ as defined in $(II.13)$ and then the noise eigenvalues will not be equal but will be absorbed with the signal eigenvalues also. In that case,

$$\mathbf{CV} = [\hat{\lambda}_1\mathbf{v}_1 \ \cdots \ \hat{\lambda}_p\mathbf{v}_p \ \ \hat{\lambda}_{p+1}\mathbf{v}_{p+1} \ \cdots \ \hat{\lambda}_N\mathbf{v}_N] \tag{IV.5}$$

where, the estimated eigenvalues are ordered as, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ \cdots \hat{\lambda}_N$. The eigenvectors corresponding to the p largest eigenvalues are called the 'signal eigenvectors' which constitute the 'signal-subspace'. All the other $(N - p)$ eigenvectors are known as the 'noise eigenvectors'. Note also that the p 'signal eigenvectors' of $\mathbf{C}$ span the subspace defined by the columns of $\mathbf{T}$ and that they are orthogonal to the 'noise subspace' eigenvectors.

V : The Proposed DFT-Based Minimum-Norm Method (D-MNM)

As a significant departure from the eigen-based approaches discussed in the previous section, this work advocates that the signal-subspace information be extracted from the DFT-of-AC matrix which can be accomplished with a single matrix multiplication. This will eliminate the need for iterative calculation of eigenvectors which is computationally intensive. The central idea behind the DFT-of-AC matrix is analyzed first.

V.a : Signal and Noise Subspace Extraction from the DFT-of-AC Matrix

Let the DFT matrix be denoted as,

$$\mathbf{D} \triangleq [\mathbf{e}_1 \ \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_N], \tag{V.1}$$

where, the elements of the k-th DFT-vector $\mathbf{e}_k$ is defined as, $\mathbf{e}_k(l) = e^{j\frac{2\pi}{N}kl}$, for $k, l = 0,1,2,\ldots, N-1$. If the frequencies ω_is are all on the DFT bins and if there is no observation noise, then in general,

$$\mathbf{f}_k \triangleq \mathbf{Ce}_k \tag{V.2a}$$

$$= \frac{1}{M}\sum_{m=1}^{M}(\mathbf{x}_m^H\mathbf{e}_k)\mathbf{x}_m, \qquad \text{using } (I.13b) \tag{V.2b}$$

$$= \frac{1}{M}\sum_{m=1}^{M}(\mathbf{a}_m^H\mathbf{T}^H\mathbf{e}_k)\mathbf{x}_m, \qquad \text{using } (I.8) \tag{V.2c}$$

$$= \frac{1}{M}\sum_{m=1}^{M}\mathbf{a}_m^H\begin{bmatrix}\mathbf{t}_1^H\mathbf{e}_k\\ \vdots \\ \mathbf{t}_p^H\mathbf{e}_k\end{bmatrix}\mathbf{x}_m. \tag{V.2d}$$

If the k-th DFT vector $\mathbf{e}_k$ corresponds to one of the ω_i frequencies,

$$\begin{aligned}\mathbf{f}_k &= \frac{1}{M}\sum_{m=1}^{M}A_{km}^*\mathbf{T}\mathbf{a}_m = \mathbf{T}\frac{1}{M}\sum_{m=1}^{M}A_{km}^*\mathbf{a}_m \\ &= \mathbf{T}\begin{bmatrix}\frac{1}{M}\sum_{m=1}^{M}A_{km}^*A_{1m}\\ \vdots \\ \frac{1}{M}\sum_{m=1}^{M}|A_{km}|^2 \\ \vdots \\ \frac{1}{M}\sum_{m=1}^{M}A_{km}^*A_{pm}\end{bmatrix} = \mathbf{T}\begin{bmatrix}\hat{\sigma}_{k1}\\ \vdots \\ \hat{\sigma}_k^2 \\ \vdots \\ \hat{\sigma}_{kp}\end{bmatrix}\end{aligned} \tag{V.3}$$

where, $\hat{\sigma}_{kl}$s denote the covariance of the complex amplitudes. Assuming the number of samples M to be large and since A_{km}s are independent random variables, $\hat{\sigma}_{A_{km},A_{lm}} \triangleq \hat{\sigma}_{kl} = \delta_{kl}\hat{\sigma}_k^2$. Hence,

$$\mathbf{f}_k \rightarrow \hat{\sigma}_k^2\mathbf{t}_k = \hat{\sigma}_k^2\mathbf{e}_k. \tag{V.4}$$

Note that the norm of $\mathbf{f}_k$ is directly proportional to the signal power, $\hat{\sigma}_k^2$, *i.e.*, this norm will be large if the signal power is significant. On the other hand, if a DFT-vector $\mathbf{e}_k$ does not correspond to any of the ω_i frequencies then due to orthogonality, $\mathbf{t}_i^H\mathbf{e}_k = 0, \forall i$. For such cases,

$$\mathbf{f}_k = \mathbf{0}. \tag{V.5}$$

For this ideal case then, the DFT-of-AC has the following decomposition,

$$\begin{aligned}\mathbf{F} &\triangleq \mathbf{C}\mathbf{D} && (V.6a)\\ &\triangleq [\mathbf{f}_1 \;\; \mathbf{f}_2 \;\; \cdots \;\; \mathbf{f}_N] && (V.6b)\\ &\rightarrow [\Lambda_1\mathbf{u}_1 \;\; \cdots \;\; \Lambda_p\mathbf{u}_p \;\; 0 \;\; \cdots \;\; 0] && (V.6c)\end{aligned}$$

where, the Λ_is and $\mathbf{u}_i$s are the lengths and unit vectors of each $\mathbf{f}_i$, respectively. Note that the unit vectors in the matrix in (V.6c) have been rearranged so that the zero/nonzero components are clustered together. Interestingly, this decomposition appears to be very similar to the usual Eigendecomposition of noiseless and ideal $\mathbf{C}$, as given by (V.2). For this ideal signal scenario again, if the DFT-of-AC is formed using the theoretical and noisy Covariance matrix of (V.3), then the decomposition has the form,

$$\begin{aligned}\mathbf{F} &= \mathbf{C}\mathbf{D} && (V.7a)\\ &= \mathbf{T}\mathbf{\Sigma}\mathbf{T}^H\mathbf{D} + \sigma_z^2\mathbf{D} && (V.7b)\\ &\rightarrow [(\Lambda_1+\sigma_z^2)\mathbf{u}_1 \;\; \cdots \;\; (\Lambda_p+\sigma_z^2)\mathbf{u}_p \;\; \sigma_z^2\mathbf{u}_{p+1} \;\; \cdots \;\; \sigma_z^2\mathbf{u}_N], && (V.7c)\end{aligned}$$

where the $\mathbf{u}_i$'s have been arranged in decreasing order of lengths. Note again that this decomposition is analogous to the one in (V.4). In this case also, the p largest-norm vectors of the DFT-of-AC matrix contain the signal subspace information.

In practice, the ω_is will not be on the DFT bins and the observations may also be noisy and hence, the decomposition in (V.6) or (V.7) will not hold. But the DFT-components ($\mathbf{f}_k$s) closer to the signal frequencies will tend to have larger norms. Hence, for the general scenario, when the observation data is noisy and the angular frequencies ω_is are arbitrarily spaced, the signal/noise subspace decomposition can be formed as :

$$\begin{aligned}\mathbf{F} &\rightarrow [\Lambda_1\mathbf{u}_1 \ \cdots \ \Lambda_p\mathbf{u}_p \ \mid \ \Lambda_{p+1}\mathbf{u}_{p+1} \ \cdots \ \Lambda_N\mathbf{u}_N] && (V.8a)\\ &\triangleq \mathbf{\Lambda}[\mathbf{U}_S \ \mid \ \mathbf{U}_N] && (V.8b)\\ &\triangleq [\mathbf{F}_S \ \mid \ \mathbf{F}_N] && (V.8c)\\ &\triangleq [\mathbf{f}_1 \ \mathbf{f}_2 \ \ldots \ \mathbf{f}_p \ \vdots \ \mathbf{f}_{p+1} \ \ldots \ \mathbf{f}_N] && (V.8d)\end{aligned}$$

where, the matrix $\mathbf{F}_S$ is formed with p number of $\mathbf{f}_i$ vectors having larger norms, $\Lambda_1 \geq \Lambda_2 \geq \cdots \geq \Lambda_N$ are the norms of the corresponding $\mathbf{f}_i$ vectors and the matrices $\mathbf{\Lambda}$, $\mathbf{U}_S$ and $\mathbf{U}_N$ are formed as,

$$\mathbf{\Lambda} \triangleq \begin{bmatrix}\Lambda_1 & & \\ & \Lambda_2 & \\ & & \ddots & \\ & & & \Lambda_p\end{bmatrix}, \quad \mathbf{U}_S \triangleq \begin{bmatrix} | & | & \cdots & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \\ | & | & \cdots & | \end{bmatrix} \quad \text{and,} \quad \mathbf{U}_N \triangleq \begin{bmatrix} | & \cdots & | \\ \mathbf{u}_{p+1} & \cdots & \mathbf{u}_N \\ | & \cdots & | \end{bmatrix}. \qquad (V.8e)$$

It may be observed again that the decomposition in (V.8) is analogous to the eigen-based counterpart in (IV.5). It may also be noted here that in case of the ideal signal cases of (V.6) and (V.7), an unit vector $\mathbf{u}_i$ corresponds to one of the DFT-vector $\mathbf{e}_k$, but in the general case of (V.8), they are linear combinations of the DFT-components close to the signal frequencies.

V.b : Incorporation of DFT-Based Signal Subspace in the Minimum-Norm Framework

The principal idea behind the Minimum-Norm method is to form an appropriate 'noise-subspace' vector $\mathbf{d}$ which is orthogonal to the 'signal-subspace' defined by $\mathbf{F}_S$. Let,

$$D(z) \triangleq \sum_{k=0}^{N-1} d_k z^{-k} \qquad (V.9)$$

be an $(N-1)$-th order z-polynomial with p zeros at, $z_i = e^{j\omega_i}$, for, $i = 1, \ldots, p$, corresponding to the DOAs. The coefficient vector is denoted as,

$$\mathbf{d} \triangleq [d_0 \ \ d_1 \ \ \cdots \ \ d_{N-1}]^T, \triangleq \begin{pmatrix}1\\ \cdots \\ \mathbf{d}'\end{pmatrix} \qquad (V.10)$$

where, $d_0 = 1$ and $\mathbf{d}'$ contains the unknown coefficients. According to the MNM philosophy [15], if $\mathbf{F}_S$ does constitute of the signal-subspace, then $\mathbf{d}$ must be orthogonal to $\mathbf{F}_S$, *i.e.*,

$$\mathbf{F}_S^H\mathbf{d} = \mathbf{0}. \qquad (V.11)$$

$\mathbf{d}$ needs to be found by solving this underdetermined set of equations which has infinite number of solutions. According to [15, 26], the solution that also minimizes the norm $||\mathbf{d}||^2$, possesses the desirable property that all its roots fall inside the unit circle. This 'minimum-norm' solution of $\mathbf{d}$ for solving (V.11) can be expressed as :

$$\mathbf{d} = \begin{bmatrix} 1 \\ -------- \\ -\ \mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{g}\end{bmatrix}, \qquad (V.12a)$$

where, $\mathbf{F}_S^H$ is partitioned as,

$$\mathbf{F}_S^H \triangleq [\mathrm{g} \mid \mathbf{G}]. \qquad (V.12b)$$

Once **d** is estimated, the p roots of $D(z)$ closest to the unit circle are used to find the DOAs. It may be recalled that in E-MNM the signal-subspace eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, as defined in (*V*.5) are used to form $\mathbf{F}_S$ [15, 26]. But in case of the proposed approach, no eigendecomposition is necessary. Post-multiplication of **C** by the DFT-matrix **D** is all that is required to extract the signal subspace in (*V*.8).

V.c : Summary of the Proposed D-MNM Algorithm

The key steps and some alternative possibilities are summarized in this Section.

V.c.1 : Algorithm Steps

1. Form the Covariance Matrix estimate using forward-backward method [15] :

$$\hat{\mathbf{C}} \triangleq \frac{1}{2M} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H + \mathbf{x}_m^b \mathbf{x}_m^{b\ H}. \qquad (V.13)$$

The 'backward' vector is defined as $\mathbf{x}_m^b \triangleq \mathbf{J}\mathbf{x}_m^*$, where, **J** denotes the permutation matrix with 1's at the cross-diagonal entries and * denotes the complex-conjugate operation.

2. Post-multiply **C** by the DFT matrix **D** to form the DFT-OF-AC matrix, $\mathbf{F} \triangleq \mathbf{CD}$.

3. Form $\mathbf{F}_S$ as in (*V*.8*c*) using the p unit vectors corresponding to the largest norms. Partition $\mathbf{F}_S$ as in (*V*.12*b*).

4. Estimate the **d** vector using (*V*.12*a*) and form the $D(z)$ polynomial using the elements of **d**.

5. Find the roots of $D(z)$. Pick the p roots closest to the unit circle to find the desired frequencies/DOAs.

V.c.2 : Alternate Possibilities

Steps 2 and 3 : Post-multiplication of the AC-matrix by a DFT-matrix has been used here because the decompositions as described in Section V appear analogous to eigendecomposition. But it is easy show that identical results can be obtained if the AC-matrix is pre-multiplied by a DFT matrix, *i.e.*, the DFT-of-AC matrix can also be formed alternately as, $\mathbf{F}_1 \triangleq \mathbf{DC}$. In that case, the largest norm row vectors of the DFT-of-AC matrix $\mathbf{F}_1$ must be used to form $\mathbf{F}_S^H$ defined in (*V*.13).

Step 4 : This step requires inversion of a matrix of dimension $(N-1) \times (N-1)$. This can be avoided by orthogonalizing the p largest norm vectors in $\mathbf{F}_S$. Let, $\mathbf{F}_S^o$ be the new 'signal-subspace' matrix with the orthonormal set of vectors which can be written in partitioned form as,

$$\mathbf{F}_S^{o\ H} \triangleq [\mathrm{g}_o \mid \mathbf{G}_o]. \qquad (V.14)$$

With these partitioned matrices, **d** can again be found in Step-4 as [15],

$$\mathbf{d} = \begin{bmatrix} 1 \\ \hline -\ \mathbf{G}_o^H \mathrm{g}_o/(1 - \mathrm{g}_o^H \mathrm{g}_o) \end{bmatrix}. \qquad (V.15)$$

It may be mentioned here that in [15], p orthonormal signal eigenvectors were used to form $\mathbf{F}_S$, whereas here $\mathbf{F}_S^o$ is formed by orthogonalizing the p largest norm vectors of the DFT-of-AC matrix.

Step 5 : This step requires rooting of the $(N-1)$-th order polynomial $D(z)$. Instead, the frequencies may also be found from the peaks of the following minimum-norm pseudo-spectrum [15, 26, 40] :

$$P_{MNM}(e^{j\omega}) \triangleq \frac{1}{|D(e^{j\omega})|^2} \tag{V.16}$$

VI : Perturbation Analysis of the D-MNM Algorithm

According to the MNM framework, the angles of arrival θ_i are extracted from the p roots of the polynomial $D(z)$ that are on or closest to the unit circle. Let the p signal zeros of $D(z)$ on or closest to the unit circle be denoted as, $z_i = e^{j\omega_i}$ which are found by rooting $D(z)$.

In practice, the polynomial $D(z)$, defined in (V.9), is formed with the coefficient vector $\mathbf{d}$ estimated using (V.12). Since $\mathbf{d}$ is a function of observations, any error in $\mathbf{d}$ would be due to deviations or noise in the observations. Error in estimated $\mathbf{d}$ would affect the estimated roots, z_i and that in turn would introduce errors in the corresponding ω_i's as well as in the θ_i's. Hence, in order to analyze the bias and MSE of the AOAs, we need to relate these errors all the way back to the error in the MNM coefficient vector $\mathbf{d}$. Hence, we begin by relating the AOA errors to signal zero errors which are then related to the coefficient errors.

VI.1 : Relationships Between the Errors in the AOA Estimates and the Signal Zeros

From (II.5) and the definition of $D(z)$ in (V.9) we know,

$$z_i = e^{j\omega_i} = e^{j\frac{2\pi d}{\lambda}\sin\theta_i} \tag{VI.1a}$$

and,

$$z_i^* = e^{-j\frac{2\pi d}{\lambda}\sin\theta_i} \tag{VI.1b}$$

Hence, we can write,

$$\begin{aligned}\frac{dz_i}{d\theta_i} &= j\frac{2\pi d}{\lambda}\cos\theta_i e^{j\frac{2\pi d}{\lambda}\sin\theta_i} \\ &\approx \frac{\Delta z_i}{\Delta\theta_i}\end{aligned} \tag{VI.2}$$

and,

$$\Delta z_i = j\frac{2\pi d}{\lambda}\cos\theta_i e^{j\frac{2\pi d}{\lambda}\sin\theta_i}\Delta\theta_i. \tag{VI.3a}$$

Similarly for z_i^*,

$$\Delta z_i^* = -j\frac{2\pi d}{\lambda}\cos\theta_i e^{-j\frac{2\pi d}{\lambda}\sin\theta_i}\Delta\theta_i. \tag{VI.3b}$$

Hence, the bias errors in AOAs have the following linear relationship with the corresponding signal root errors,

$$E(\Delta\theta_i) = -j\frac{\lambda}{2\pi d\cos\theta_i}e^{-j\frac{2\pi d}{\lambda}\sin\theta_i}E(\Delta z_i). \tag{VI.4}$$

where 'E( )' is used to denote the expectation operator. Furthermore, using (VI.3) the MSE of the AOAs can be shown to be related to the signal-root MSEs as,

$$E(|\Delta\theta_i|^2) = (\frac{\lambda}{2\pi d\cos\theta_i})^2 E(|\Delta z_i|^2). \tag{VI.5}$$

Equation (VI.4) and (VI.5) show that the bias and mean squared error in the AOA estimate are linear to those in the signal zeros. Next, the bias and MSE of signal zeros are related to the coefficient errors.

VI.2 : Effect of Coefficient Error on the Bias and Mean-Squared Error of the Zeros

The effect of errors in the coefficients d_k on the zeros of $D(z)$ is well known [11] and is given by (note that, $d_0 = 1$),

$$\frac{\partial z_i}{\partial d_k} = \frac{-z_i^{-(k-1)}}{\prod_{l=1,l\neq i}^{N-1} (1 - z_l z_i^{-1})}; \qquad k = 1, 2, \ldots, N-1. \tag{VI.6}$$

Using (VI.6), the total error in z_i due to error in all coefficients is given by,

$$\begin{aligned}\Delta z_i &\approx \sum_{k=1}^{N-1} \frac{\partial z_i}{\partial d_k} \Delta d_k \\ &= \frac{-1}{\prod_{l=1,l\neq i}^{N-1} (1 - z_l z_i^{-1})} [1 \ z_i^{-1} \ z_i^{-2} \ \ldots \ z_i^{-(N-2)}] \Delta \mathbf{d}' \\ &= \frac{-\sqrt{N-1}}{\prod_{l=1,l\neq i}^{N-1} (1 - z_l z_i^{-1})} \mathbf{V}^H(e^{j\omega_i}) \Delta \mathbf{d}' \end{aligned} \tag{VI.7a}$$

where, $\mathbf{d}'$ is defined in (V.10) and

$$\mathbf{V}(e^{j\omega_i}) \triangleq \mathbf{V}(z_i) \triangleq \frac{1}{\sqrt{N-1}} [1 \ e^{j\omega_i} \ \ldots \ e^{j(N-2)\omega_i}]^T. \tag{VI.7b}$$

Hence, the bias error of the estimated signal root is given by,

$$E(\Delta z_i) = \frac{-\sqrt{N-1}}{\prod_{l=1,l\neq i}^{N-1} (1 - z_l z_i^{-1})} \mathbf{V}^H(e^{j\omega_i}) E(\Delta \mathbf{d}') \tag{VI.8}$$

The mean squared error is given by,

$$E(|\Delta z_i|^2) = \frac{N-1}{\prod_{l=1,l\neq i}^{N-1} |1 - z_l z_i^{-1}|^2} \mathbf{V}^H(e^{j\omega_i}) E(\Delta \mathbf{d}' \Delta \mathbf{d}'^H) \mathbf{V}(e^{j\omega_i}) \tag{VI.9a}$$

$$\triangleq S \ \mathbf{V}^H(e^{j\omega_i}) E(\Delta \mathbf{d}' \Delta \mathbf{d}'^H) \mathbf{V}(e^{j\omega_i}) \tag{VI.9b}$$

where,

$$S \triangleq \frac{N-1}{\prod_{l=1,l\neq i}^{N-1} |1 - z_l z_i^{-1}|^2} \tag{VI.9c}$$

denotes the sensitivity of the parameter set and is a measure of the effect of errors in the parameter vector **d** on the signal zeros. Equations (VI.8) shows that the bias in the signal-zeros is linearly related to the bias in the coefficient vector and equation (VI.9) provides the relationship between the MSE of signal roots and the coefficient error covariance matrix. The expressions for $E(\Delta \mathbf{d}')$ and $E(\Delta \mathbf{d}' \Delta \mathbf{d}'^H)$ are derived next.

VI.3 : Coefficient Error due to the D-MNM Method

The analysis in this part would rely on the assumption made at the outset; the observation data consists of p complex sinusoids in additive white Gaussian noise $z_n(m)$. Rewriting explicitly the observation matrix defined in (II.12),

$$\mathbf{X} \triangleq [\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_M]$$
$$\triangleq \begin{pmatrix} x_1(0) & x_2(0) & \ldots & x_M(0) \\ x_1(1) & x_2(1) & \ldots & x_M(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N-1) & x_2(N-1) & \ldots & x_M(N-1) \end{pmatrix} \qquad (VI.10a)$$

where,

$$x_m(n) \triangleq \sum_{i=1}^{p} A_{im} e^{j\omega_i n} + z_m(n). \qquad (VI.10b)$$

Note that the phase $\phi_m(i)$ in A_{im} which is defined in (VI.6) is assumed to be uniformly distributed. In addition, the theoretical autocorrelation matrix $\mathbf{C}$ in (IV.3) can be written explicitly as,

$$\mathbf{C} \triangleq E(\mathbf{X}\mathbf{X}^H)$$
$$= \mathbf{T\Sigma T}^H + \sigma_z^2\mathbf{I}$$
$$= \begin{pmatrix} E(x(0)x^*(0)) & E(x(0)x^*(1)) & \ldots & E(x(0)x^*(N-1)) \\ E(x(1)x^*(0)) & E(x(1)x^*(1)) & \ldots & E(x(1)x^*(N-1)) \\ \vdots & \vdots & \ddots & \vdots \\ E(x(N-1)x^*(0)) & E(x(N-1)x^*(1)) & \ldots & E(x(N-1)x^*(N-1)) \end{pmatrix}$$
$$= \begin{pmatrix} C(0) & C(-1) & C(-2) & \ldots & C(-(N-1)) \\ C(1) & C(0) & C(-1) & \ldots & C(-(N-2)) \\ C(2) & C(1) & C(0) & \ldots & C(-(N-3)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C(N-1) & C(N-2) & C(N-3) & \ldots & C(0) \end{pmatrix} \qquad (VI.11)$$

where,

$$C(m) = \begin{cases} \sum_{i=1}^{p} \ |\sigma_i^2|^2 + \sigma_z^2, & \text{if } m = 0; \\ \sum_{i=1}^{p} \ |\sigma_i^2|^2 e^{j\omega_i m}, & \text{if } m \neq 0. \end{cases} \qquad (VI.12)$$

From Section V.a we know that the D-MNM method uses the following decomposition,

$$\mathbf{F} \triangleq \mathbf{CD} \qquad (VI.13a)$$

$$\rightarrow [\mathbf{F}_S \ \vdots \ \mathbf{F}_N] \qquad (VI.13b)$$

$$\triangleq [\mathbf{f}_1 \ \mathbf{f}_2 \ \ldots \ \mathbf{f}_p \ \vdots \ \mathbf{f}_{p+1} \ \ldots, \mathbf{f}_N] \qquad (VI.13c)$$

$$= [\mathbf{Ce}_1 \ \mathbf{Ce}_2 \ \ldots \ \mathbf{Ce}_p \ \vdots \ \mathbf{Ce}_{p+1} \ \ldots \ \mathbf{Ce}_N] \qquad (VI.13d)$$

$$= [\mathbf{CD}_S \ \vdots \ \mathbf{CD}_N] \qquad (VI.13e)$$

where p is the assumed number of signal sources. $\mathbf{D}_S$ contains the signal subspace DFT-vectors $\mathbf{e}_i$ which correspond to the largest norm $\mathbf{f}_i$ vectors. In D-MNM, the vector $\mathbf{d}$ is entirely in the noise subspace and must be orthogonal to the signal subspace $\mathbf{F}_S$, i.e., repeating from (V.11)

$$\mathbf{F}_S^H\mathbf{d} = \mathbf{0}, \qquad (VI.14a)$$

or, using (VI.13),

$$\mathbf{D}_S^H\mathbf{C}\mathbf{d} = \mathbf{0}, \qquad (VI.14b)$$

Using (V.10) and (V.12b),

$$\mathbf{G}\mathbf{d}' = -\mathbf{g} \qquad (VI.15)$$

and the error expression can be written as,

$$\Delta(\mathbf{G}\mathbf{d}') = -\Delta\mathbf{g}. \qquad (VI.16)$$

Using chain-rule,

$$\Delta\mathbf{G}\mathbf{d}' + \mathbf{G}\Delta\mathbf{d}' = -\Delta\mathbf{g}. \qquad (VI.17)$$

The pseudo-inverse solution for the coefficient error is,

$$\begin{aligned} \Delta\mathbf{d}' &= -\mathbf{G}^{\#}(\Delta\mathbf{G}\mathbf{d}' + \Delta\mathbf{g}) \\ &= -\mathbf{G}^{\#}(\Delta\mathbf{g} \vdots \Delta\mathbf{G}) \begin{pmatrix} 1 \\ \cdots \\ \mathbf{d}' \end{pmatrix} \\ &= -\mathbf{G}^{\#}\Delta\mathbf{F}_S^H\mathbf{d} \qquad (VI.18) \end{aligned}$$

where $\mathbf{G}^{\#}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{G}$. Equation (VI.18) shows that the error in the coefficient vector depends directly on the error of the signal subspace obtained by using the D-MNM method. The coefficient error in (VI.18) can now be utilized in the expressions for bias and mean squared error of the zeros which were derived in (VI.8) and (VI.9), respectively. However, according to (VI.8) and (VI.9), in order to obtain $E(\Delta z_i)$ and $E(|\Delta z_i|^2)$ we need expressions for $E(\Delta')$ and $E(\Delta'\Delta'^H)$, which are derived in the next two sections.

VI.4 : Bias in the Signal Zeros

First let us find the mean of the coefficient error,

$$\begin{aligned} E(\Delta\mathbf{d}') &= -\mathbf{G}^{\#}E(\Delta\mathbf{F}_S^H)\mathbf{d} \\ &= -\mathbf{G}^{\#}E(\mathbf{D}_S^H\Delta\mathbf{C}^H)\mathbf{d}; \qquad \text{using}(VI.13) \\ &= -\mathbf{G}^{\#}E(\mathbf{D}_S^H\hat{\mathbf{C}})\mathbf{d} + \mathbf{G}^{\#}\mathbf{D}_S^H\mathbf{C}\mathbf{d}; \qquad \text{assuming}\Delta\mathbf{C} = \hat{\mathbf{C}} - \mathbf{C} \\ &= -\mathbf{G}^{\#}\mathbf{D}_S^H E(\hat{\mathbf{C}})\mathbf{d} \qquad \text{using}(VI.14) \\ &\approx -\mathbf{G}^{\#}\mathbf{D}_S^H\mathbf{C}\mathbf{d} \qquad \text{if covariance is unbiased} \\ &= \mathbf{0}; \qquad \textbf{using(VI.14)}. \\ & (VI.19) \end{aligned}$$

Substituting (VI.19a) in (VI.8) results in,

$$E(\Delta\mathbf{z}_i) \approx \mathbf{0}. \qquad (VI.20)$$

This expression implies that the bias in the estimate obtained by using the D-MNM method can be expected to be quite small. This fact was observed in simulations also.

VI.5 : Mean Squared Error in the Signal Zeros

For obtaining an estimate of the mean squared error in the zeros, an expression is needed for $E(\Delta\mathbf{d}'\Delta\mathbf{d}'^H)$ which appears in (VI.9). Starting with (VI.18), we have,

$$\begin{aligned}
&E(\Delta\mathbf{d}'\Delta\mathbf{d}'^H)\\
&= \mathbf{G}^{\#}E(\Delta\mathbf{F}_S^H\mathbf{d}\mathbf{d}^H\Delta\mathbf{F}_S)(\mathbf{G}^H)^{\#}\\
&= \mathbf{G}^{\#}\mathbf{D}_S^H E(\Delta\mathbf{C}^H\mathbf{d}\mathbf{d}^H\Delta\mathbf{C})\mathbf{D}_S(\mathbf{G}^H)^{\#}
\end{aligned}\tag{VI.21}$$

where,

$$\begin{aligned}
&E(\Delta\mathbf{C}^H\mathbf{d}\mathbf{d}^H\Delta\mathbf{C})\\
&= E((\hat{\mathbf{C}}^H - \mathbf{C}^H)\mathbf{d}\mathbf{d}^H(\hat{\mathbf{C}} - \mathbf{C}))\\
&= E(\hat{\mathbf{C}}^H\mathbf{d}\mathbf{d}^H\hat{\mathbf{C}}) - E(\mathbf{C}^H\mathbf{d}\mathbf{d}^H\hat{\mathbf{C}}) - E(\hat{\mathbf{C}}^H\mathbf{d}\mathbf{d}^H\mathbf{C}) + E(\mathbf{C}^H\mathbf{d}\mathbf{d}^H\mathbf{C})\\
&= E(\hat{\mathbf{C}}^H\mathbf{d}\mathbf{d}^H\hat{\mathbf{C}}) - \mathbf{C}^H\mathbf{d}\mathbf{d}^H\mathbf{C}
\end{aligned}\tag{VI.22}$$

and from (II.13), we know,

$$\begin{aligned}
&E(\hat{\mathbf{C}}^H\mathbf{d}\mathbf{d}^H\hat{\mathbf{C}})\\
&= E((\frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i\mathbf{x}_i^H)\mathbf{d}\mathbf{d}^H(\frac{1}{M}\sum_{j=1}^{M}\mathbf{x}_j\mathbf{x}_j^H))\\
&= \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} E(\mathbf{x}_i\mathbf{x}_i^H\mathbf{d}\mathbf{d}^H\mathbf{x}_j\mathbf{x}_j^H)
\end{aligned}\tag{VI.23}$$

Since this expression involves fourth moments of the process, its computation would be difficult in general. However, for a Gaussian random process all higher-order moments can be expressed in term of first and second moments. In particular, if v_1, v_2, v_3, and v_4 are complex Gaussian random variables, it is known that [40],

$$E(v_1v_2^*v_3v_4^*) = E(v_1v_2^*)E(v_3v_4^*) + E(v_1v_4^*)E(v_3v_2^*)\tag{VI.24}$$

Applying (VI.24) to (VI.23) leads to the following expression,

$$\begin{aligned}
&E(\hat{\mathbf{C}}^H\mathbf{d}\mathbf{d}^H\hat{\mathbf{C}})\\
&= \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}E\Big\{\begin{pmatrix} x_i(1)\\ \vdots\\ x_i(N)\end{pmatrix}(x_i^*(0)\ldots)\begin{pmatrix} d_1\\ \vdots\\ d_N\end{pmatrix}(d_1^*\ldots)\begin{pmatrix} x_j(0)\\ \vdots\\ x_j(N-1)\end{pmatrix}(x_j^*(0)\ldots x_j^*(N-1))\Big\}\\
&= \frac{1}{M^2}\sum_{ijgk} d_kd_g^*\begin{pmatrix} E(x_i(0)x_i^*(k)x_j(g)x_j^*(0)) & \ldots & E(x_i(0)x_i^*(k)x_j(g)x_j^*(N-1))\\ \vdots & \ddots & \vdots\\ E(x_i(N-1)x_i^*(k)x_j(g)x_j^*(0)) & \ldots & E(x_i(N-1)x_i^*(k)x_j(g)x_j^*(N-1))\end{pmatrix}\\
&= \frac{1}{M^2}\sum_{ijgk} d_kd_g^*\begin{pmatrix} E(x_i(0)x_i^*(k))E(x_j(g)x_j^*(0)) & \ldots & E(x_i(0)x_i^*(k))E(x_j(g)x_j^*(N-1))\\ \vdots & \ddots & \vdots\\ E(x_i(N-1)x_i^*(k))E(x_j(g)x_j^*(0)) & \ldots & E(x_i(N-1)x_i^*(k))E(x_j(g)x_j^*(N-1))\end{pmatrix}\\
&\quad + \frac{1}{M^2}\sum_{ijgk} d_kd_g^*\begin{pmatrix} E(x_i(0)x_j^*(0))E(x_j(g)x_i^*(k)) & \ldots & E(x_i(0)x_j^*(N-1))E(x_j(g)x_i^*(k))\\ \vdots & \ddots & \vdots\\ E(x_i(N-1)x_j^*(0))E(x_j(g)x_i^*(k)) & \ldots & E(x_i(N-1)x_j^*(N-1))E(x_j(g)x_i^*(k))\end{pmatrix}\\
&= \mathbf{C}^H\mathbf{d}\mathbf{d}^H\mathbf{C} + \frac{1}{M^2}\sum_{ijgk} d_kd_g^*E(x_j(g)x_i^*(k))\begin{pmatrix} E(x_i(0)x_j^*(0)) & \ldots & E(x_i(0)x_j^*(N-1))\\ \vdots & \ddots & \vdots\\ E(x_i(N-1)x_j^*(0)) & \ldots & E(x_i(N-1)x_j^*(N-1))\end{pmatrix}
\end{aligned}\tag{VI.25}$$

where,

$$\sum_{ijgk} \triangleq \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{g=1}^{N}\sum_{k=1}^{N}$$

Substituting (VI.25) in (VI.22) results in,

$$E(\Delta\mathbf{C}^H\mathbf{d}\mathbf{d}^H\Delta\mathbf{C})$$

$$= \frac{1}{M^2}\sum_{ijgk} d_k d_g^* E(x_j(g)x_i^*(k)) \begin{pmatrix} E(x_i(0)x_j^*(0)) & \dots & E(x_i(0)x_j^*(N-1)) \\ \vdots & \ddots & \vdots \\ E(x_i(N-1)x_j^*(0)) & \dots & E(x_i(N-1)x_j^*(N-1)) \end{pmatrix} \qquad (VI.26)$$

Further, if the noise is white and the complex amplitudes of the signal are uncorrelated, then if $i \neq j$, $E(x_j(g)x_i^*(k)) = 0$. In (VI.26), among M^2 possible i, j combinations, there are only M terms for which, $i = j$. Retaining only these terms we have, (assuming stationarity, the dependence on i $(=j)$ has been suppressed)

$$E(\Delta\mathbf{C}^H\mathbf{d}\mathbf{d}^H\Delta\mathbf{C})$$

$$= \frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N} d_k d_g^* E(x(g)x^*(k)) \begin{pmatrix} E(x(0)x^*(0)) & \dots & E(x(0)x^*(N-1) \\ \vdots & \ddots & \vdots \\ E(x(N-1)x^*(0)) & \dots & E(x(N-1)x^*(N-1)) \end{pmatrix}$$

$$= \frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N} d_k d_g^* E(x(g)x^*(k)) \ \mathbf{C} \qquad (VI.27a)$$

$$= \frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N} d_k d_g^* (\sum_{i=1}^{p} |a_i|^2 e^{j\omega_i(g-k)} \ + \ \sigma_z^2\delta_{gk}) \ \mathbf{C} \qquad (VI.27b)$$

in the above, δ_{gk} is the Kronecker delta and hence,

$$E(x(g)x^*(k)) \ = \ \begin{cases} C(0) \ = \ \sum_{i=1}^{p} |a_i|^2 \ + \ \sigma_z^2, & \text{if } g \ = \ k, \\ C(g-k) \ = \ \sum_{i=1}^{p} |a_i|^2 e^{j\omega_i(g-k)}, & \text{if } g \neq k. \end{cases} \qquad (VI.27c)$$

Substituting (VI.27b) in (VI.21),

$$E(\Delta\mathbf{d}'\Delta\mathbf{d}'^H) \ = \ \mathbf{G}^{\#}\mathbf{D}_S^H\Big(\frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N} d_k d_g^* (\sum_{i=1}^{p} |a_i|^2 e^{j\omega_i(g-k)} \ + \ \sigma_z^2\delta_{gk})\mathbf{C}\Big)\mathbf{D}_S(\mathbf{G}^H)^{\#}. \qquad (VI.28)$$

Using this in the expression of the MSE of signal zeros in (VI.9b),

$$E(|\Delta z_i|^2) \ = \ S\mathbf{V}^H(e^{j\omega_i})\mathbf{G}^{\#}\mathbf{D}_S^H\Big(\frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N} d_k d_g^* (\sum_{i=1}^{p} |a_i|^2 e^{j\omega_i(g-k)} \ + \ \sigma_z^2\delta_{gk})\mathbf{C}\Big)\mathbf{D}_S(\mathbf{G}^H)^{\#}\mathbf{V}(e^{j\omega_i}) \quad (VI.29)$$

Now we are ready to obtain the final expression for the bias and MSE of the DOA estimates.

VI.6 : Bias in the AOA Estimates

Using the developments in (VI.19)-(VI.20) into the AOA bias equation (VI.4),

$$E(\Delta\theta_i) \ = \ \mathbf{0}. \qquad (VI.30)$$

VI.7 : Mean Squared Error in the AOA Estimates

Using (VI.29) in (VI.5), we finally obtain the MSE expression of AOAs,

$$E(|\Delta\theta_i|^2) \;=\; S(\frac{\lambda}{2\pi d\cos\theta_i})^2\mathbf{V}^H(e^{j\omega_i})\mathbf{G}^{\#}\mathbf{D}_S^H\Big(\frac{1}{M}\sum_{g=1}^{N}\sum_{k=1}^{N}d_k d_g^*(\sum_{i=1}^{p}|a_i|^2e^{j\omega_i(g-k)} \;+\; \sigma_z^2\delta_{gk})\mathbf{C}\Big)\mathbf{D}_S(\mathbf{G}^H)^{\#}\mathbf{V}(e^{j\omega_i}) \qquad (VI.31)$$

This expression explicitly shows that the mean squared error in the AOAs not only depends on the parameter sensitivities but also on the specific structure in the D-MNM method. It should be mentioned here some of developments presented here are similar to the work in [25]. However, the results in [25] depends heavily on statistical properties of eigenvalues/eigenvectors which are neither applicable not appropriate in the present case. Hence, in the development above the moments of the data and covariance have been used directly. Simulation studies with similar data sets used in [25] verify close match between the theory and simulation results.

VII : Simulation Results

In the first two examples, the performance of the proposed D-MNM algorithm is compared with the performance of some of the existing well-known algorithms using Monte-Carlo studies. In Example 3, the theoretically derived MSE formulae are verified by comparing theory and simulation studies. The theoretical performance is also compared with some of the eigen-based counterparts.

VII.a : DOA Estimation

Simulation 1 : *Two Closely-Spaced Targets of Equal Powers* [15, 38, 39]

Planewaves from $p = 2$ sources with $\theta_1 = 18°$ and $\theta_2 = 22°$ incident on N=8 sensors were modeled as in [15, 16, 18]. The number of snapshots, M=10. Fig. 1 shows the norms of the $\mathbf{f}_i$ vectors for 20 trials at 20dB SNR. The two largest Λ_is always appear to be more significant than all the smaller ones. Figures 2a and 2b show the roots of $D(z)$ for 50 independent realizations using D-MNM and E-MNM, respectively. The figures show that the roots in both cases are at almost same locations. Table-1 compares E-MNM and D-MNM in terms of the bias and RMS values with 200 independent trials at different SNR values. The results clearly indicate that the performance of D-MNM is quite close to that of E-MNM, though no Eigendecomposition was required in this case. In fact, D-MNM was found to be somewhat more robust (in terms of successful trials) at low SNR ranges.

VII.b : Frequency Estimation

In this Section, the proposed algorithm is compared with the well-known Tufts-Kumaresan (TK) method [17, 42] and MUSIC method [1, 30] via simulations.

Simulation 2 : *Comparison of High-Resolution Performance and Threshold Enhancement*

The simulation data is generated using the formula [42] :

$$y(n) \;=\; a_1e^{j2\pi(0.5)n+j\frac{\pi}{4}} \;+\; a_2e^{j2\pi(0.52)n} \;+\; w(n), \qquad \text{for, } n = 0,\ 1,\ \ldots,\ M-1 \qquad (VII.1)$$

where, $w(n)$ is complex white Gaussian noise with variance σ_w^2. The number of data samples used is, M=25. This data set has been widely used in the literature for studying the performance of various methods. For this data set, it has been shown in [42] that the TK method performs best when high-order $(L \times L)$ covariance matrix with $L = 18$ is used with forward-backward covariance matrix [42]. Five hundred independent noise realizations were used to compare the performance of the proposed method with that of TK method and MUSIC. The mean values for three cases at different SNR values are displayed in Fig. 4. The RMSE results are shown in Fig. 5 along with CR Bound for the frequency at $f_1 = 0.52Hz$. The bias and RMSE at different SNR values are also tabulated in Table 2. Clearly, the proposed method extends the performance threshold closer to the CR bound. Hence the performance of the proposed method approaches that of the Maximum-Likelihood method more closely, although with considerably less computational complexity.

VII.b : Perturbation Analysis of D-MNM

In this Section, the theoretical formulae derived for MSE and Bias for the DOA estimation using the proposed DFT-based algorithm are verified by comparing theoretical and simulation results.

Simulation 3 : *Perturbation Analysis of AOAs with Small Number of Sensors*

The problem scenario is identical to as described in Simulation 1 for AOA estimation. For this case, $N = 8, M = 100$, and the AOAs are 18° and 22°. In these simulations the theoretical formulas for mean-squared error, as derived in (66) and (67), are compared with the performance using simulated data. The results for various signal-to-noise ratios (SNR) are shown in Fig. 6 which shows the result corresponding to the AOA of 18° using all of the 200 independent trials.

Note that data set considered in Simulation 3 were also used in [25] for perturbation analysis of the eigen-based MNM and MUSIC. But it appears that in [25], only the results with successful trials were plotted. Fig. 6 indicates that for this example, the D-MNM method appears to have smaller squared error than E-MNM, especially at low SNR. This was found to be the case for all trials. In fact, the success rate was higher for D-MNM when compared with E-MNM. Following the trend in the eigen-based cases, Root-MUSIC fared a little better (except at low SNR), but it may be noted that in case of Root-MUSIC the polynomial to be rooted has twice the order than either D-MNM or E-MNM. Finally, it may be emphasized here that the theoretical predictions based on formulas derived in this paper were found to be quite close to those obtained by computer simulations.

VIII : Analysis, Discussion and Directions on Further Research

The results presented so far are quite intriguing and can may possibly have some important consequences on simplifying the present practice of frequency/DOA estimation. The proposed approach of forming signal-subspace using DFT without any eigendecomposition also opens up whole new avenues for further research and, at the same time, poses some unanswered questions. Furthermore, it may be possible to extend and incorporate similar ideas in other closely related problems or to develop more simplified algorithms. Clearly, the major advantage of the proposed approach is that all the signal-subspaces are obtained with a single matrix multiplication. This step may be performed using FFT which is very efficient for hardware and software implementation. In the following, some analysis as well as some possible directions for further research are briefly discussed.

1. **Reduced Computational Complexity and Usefulness in High Sampling-Rate Problems** : The major significance of D-MNM is that its high-resolution capability does not rely on any iterative method or eigendecomposition which is also computed iteratively. The lower computational complexity of D-MNM should be attractive in any general frequency/DOA estimation scenario. But the usefulness of the proposed method should be specially significant in those applications where traditional high-resolution methods are yet to make much inroads due mainly to extremely high sampling rate requirements. Specifically, in Electronic

Warfare (EW) applications, the signals usually operate in the GHz range but real-time, high-resolution capability is a necessity [41]. Currently no EW receiver processes signals entirely in digital. The proposed DFT-based MNM with its low computational complexity, is expected to provide the desired high-resolution capability to future digital EW receivers.

2. **Signal-Subspace Information from the Autocorrelation Matrix Only** : The strength of the Minimum-Norm framework as a high-resolution method really comes from its ability to form the 'noise-subspace' vector $\mathbf{d}$ by exploiting the orthogonality property in (V.11). It appears that as long as $\mathbf{F}_S$ has some component of the signal-subspace $\mathbf{T}$, the solution of (V.11) would retain its high-resolution capability. The DFT-of-AC is an appropriate candidate to produce $\mathbf{F}_S$ because it is a linear combination of the signal-vectors in $\mathbf{T}$. This can be seen by rewriting the DFT-of-AC matrix,

$$\mathbf{F} = \mathbf{CD} = \mathbf{T}\left[\frac{1}{M}\sum_{m=1}^{M}\mathbf{a}_m(\mathbf{x}_m^H\mathbf{D})\right]. \qquad (VIII.1)$$

In fact, the AC matrix itself is also a possible candidate for obtaining the 'signal-subspace' $\mathbf{F}_S$, because it can be expressed as a linear combination of the signal-vectors in $\mathbf{T}$,

$$\mathbf{C} = \mathbf{T}\left[\frac{1}{M}\sum_{m=1}^{M}\mathbf{a}_m\mathbf{x}_m^H\right]. \qquad (VIII.2)$$

Theoretically, the norm of each vector in ideal $\mathbf{C}$ should be equal but with noisy $\hat{\mathbf{C}}$, the norms of some of the vectors may be reduced while for other vectors, the norms may be more than the nominal value. Hence, the ideal choice would be to pick the vectors with norms in the middle range. Not surprisingly, when $\mathbf{F}_S$ is formed in this manner with p vectors of the estimated $\mathbf{C}$, MNM again demonstrated high-resolution capability in simulations (not included). This simpler procedure to obtain 'signal-subspace' information needs to be studied further. But it must be stated that D-MNM performs better at low SNR because the DFT operation accentuates the signal-subspace, as discussed next.

3. **Asymptotic Analysis of the DFT-based Signal Subspace for Arbitrary DOA/Frequency** : For ideal noise-free observations if the frequencies are not on the DFT bins, the DFT-of-AC operation can be expressed as :

$$\begin{aligned}\mathbf{F} &= \mathbf{CD} && (VIII.3a)\\ &= \mathbf{T\Sigma T}^H\mathbf{D} && (VIII.3b)\\ &= \mathbf{T\Sigma}\begin{bmatrix}\mathbf{t}_1^H\mathbf{D}\\ \mathbf{t}_2^H\mathbf{D}\\ \vdots\\ \mathbf{t}_p^H\mathbf{D}\end{bmatrix}. && (VIII.3c)\end{aligned}$$

Consider the matrix at right. Each of the $\mathbf{t}_i^H\mathbf{D}$ vectors are complex valued DFT of a sequence of a complex sinusoid. The magnitude of each row vector, $\mathbf{t}_i^H\mathbf{D}$ has a *Sinc* envelope with a peak occurring at the column corresponding to the bin location closest to the frequency ω_i. For infinite aperture with $N \rightarrow \infty$, *i.e.*, for large number of sensors, each row vector peaks at ω_i and the other elements of that row approaches zero. The same will be the case for each of the other row vectors also. Hence, asymptotically, the DFT-of-AC operation again produces p largest norm vectors at the true signal frequencies. The asymptotic analysis for the noisy case as defined by (V.7) would also provide similar results. For finite N, because of the Sinc weighting, the largest norm vectors will also have contributions from some other $\mathbf{t}_i$ vectors in the $\mathbf{T}$. But those components also contain signal-subspace information which is orthogonal to $\mathbf{d}$ and hence useful for obtaining the minimum-norm vector $\mathbf{d}$.

4. **Estimation of the Parameters of Damped Sinusoids in Noise** : Many eigen-based methods have been successfully utilized in estimating the unknown parameters of damped sinusoids from noisy observations [5, 14]. It appears that with some simple modifications the proposed DFT-based approach could also be used for the same purpose. The advantage would again be that no eigendecomposition but the performance will be comparable.

5. **Largest Norms vs. Peaks** : In all the simulations presented here, the signal subspaces have been formed by selecting the p unit-vectors having largest norms. But the ideal solution may be to pick the unit vectors corresponding to the p largest peaks (having smaller norm vectors on both adjacent bins). This may eliminate any possibility of picking multiple vectors from the vicinity of strong signals. It should be emphasized though that largest norm criteria has worked quite well so far, as demonstrated by a large number of simulations. But this aspect certainly needs further analysis.

6. **Zero padding** : In classical spectral estimation, Periodogram relies on DFT/FFT, but it is often necessary to extend (or, pad) the available data with zeros so that interpolated values between available bins can be calculated. Zero-padding is also used to extend data-lengths to powers of two such that the computational efficiency of the FFT can be taken advantage of. In the simulation studies, no zero-padding had been incorporated so far. It is not quite apparent whether the zero-padding should be done directly to the data or to the covariance estimates and this aspect needs further study. It would also be necessary to study the possible effects on the signal-subspace produced by the DFT-of-AC operation after zero-padding is introduced.

7. **Windowing** : In classical spectral estimation, in order to avoid sudden discontinuities, the observed data is often weighted (or tapered at both ends) by non-rectangular window which tends to enhance the 'dynamic range' at the cost of 'resolution'. In the simulation results presented here, no windowing has been used. But windowing is known to be highly effective in locating weak frequency components which tend to get submerged by the sidelobes of strong components. Though it is believed that that orthogonality property in (V.11) is the main contributing factor for the high-resolution capability of D-MNM, it would certainly be interesting to study what effects windowing might have on the performance of D-MNM.

8. **Use of DFT-Based Signal-Subspace in other Eigen-Based Methods** : Other than the Minimum-Norm Method covered in this paper, there is a large body of work where some form of eigendecomposition is utilized to estimate DOA/Frequencies [1, 2, 8, 10, 13-15, 19, 22, 23, 30, 32, 36, 42, 43, 44-47, 49]. Among the more important results are, MUSIC [30], SVD [15, 42] and ESPRIT [22]. It is quite possible that the proposed DFT-based signal-subspace may be incorporated with some of these existing eigendecomposition based methods, in order to implement those methods without eigendecomposition. Clearly, the proposed approach can be used to implement MUSIC, except that the noise subspace $\mathbf{F}_N$ defined in (V.8c) would have to be utilized. Also, the left and right eigenvectors of the SVD of a data matrix are actually the eigenvectors of correlation matrices. Hence, it appears that some of the SVD-based approaches may also be modified to incorporate DFT-based signal/noise subspaces. Care should be taken about the choice of either the left or right signal-spaces, because both may not contain signal information. The case is not so apparent for those methods which use generalized eigendecomposition [22, 36, 45]. As part of this paper, some of these possibilities will be further investigated.

9. **DFT-Prony** : There has been some recent interest in implementing the Prony's algorithm in the Frequency-Domain [29]. Clearly, the signal-vectors in $\mathbf{F}_S$ can be treated as multiple time-series to form a $(p+1) \times (p+1)$ covariance matrix (using forward-backward approach) and then the p-th order Prony's polynomial can be estimated. Based on preliminary simulations (not included), this approach appears to be simplest of all existing methods with moderately good high-resolution performance. The performance of DFT-Prony is much better than that of the standard Prony's method because the DFT-based signal subspace is cleaned-up though without any eigendecomposition. These ideas will be further studied as part of this paper.

10. **Two-Dimensional Frequency-Wavenumber Estimation** : In some array processing scenarios, both the DOAs (related to wavenumbers) and the center frequencies need to be estimated. Many existing 1-D eigen-based methods have been extended to 2-D to address this problem. It appears that the DFT-of-AC vectors can be formed in both domains and two $D(z)$ polynomials can be be formed to estimate the the frequencies and DOAs separately. Incorporation of the DFT-based signal-spaces for 2D frequency estimation will be further investigated as part of this paper.

11. **Hardware Implementation** : Perhaps the most important and useful practical impact of the proposed method would be in the area of hardware implementation for high-resolution Direction-of-Arrival or frequency estimation. All the currently available methods with good-enough high-resolution capability, rely on some form of iterative optimization or iterative computation of eigenvectors. In contrast, all that the proposed approach requires to form the 'signal-subspace' is a single matrix multiplication. Furthermore, the matrix to be multiplied is a DFT matrix and it has special structures so that FFT based processing may be utilized to further reduce the computational burden. Hence, one of the major goals of the proposed work would be to devise appropriate strategies to design, develop and, if possible, fabricate VLSI hardware for high-resolution DOA/Frequency estimation.

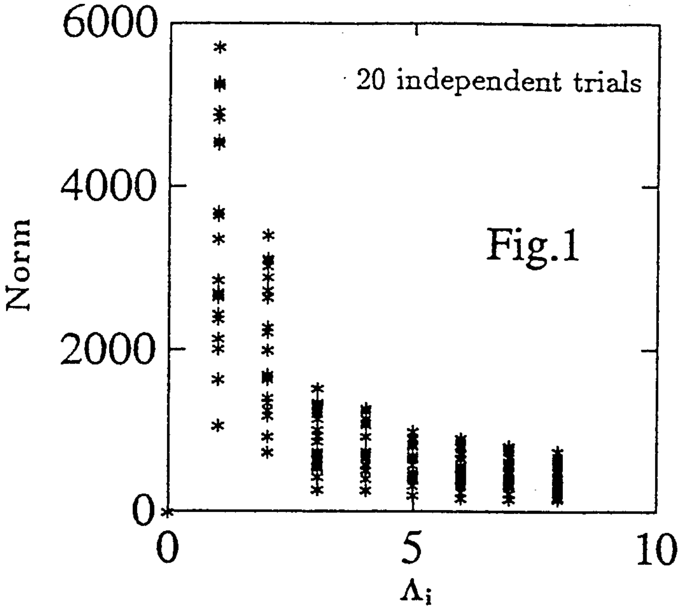


Fig.1. Norms of the DFT-of-AC vectors

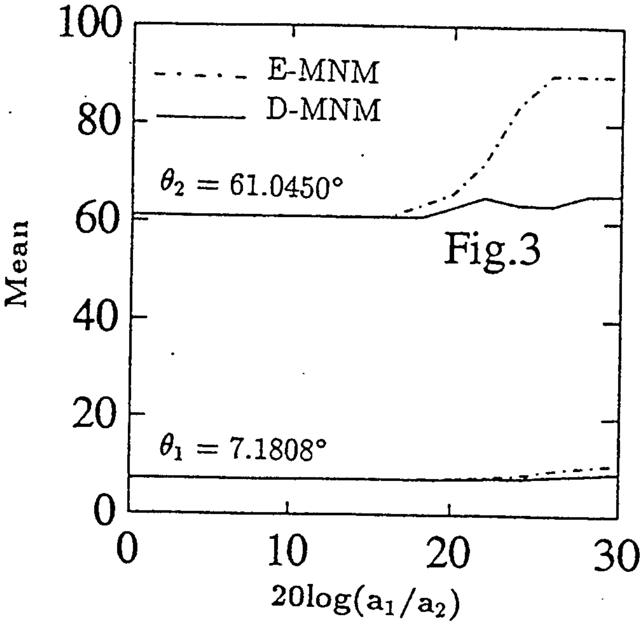


Fig.3. Means of θ_1 and θ_2 for 50 independent trials

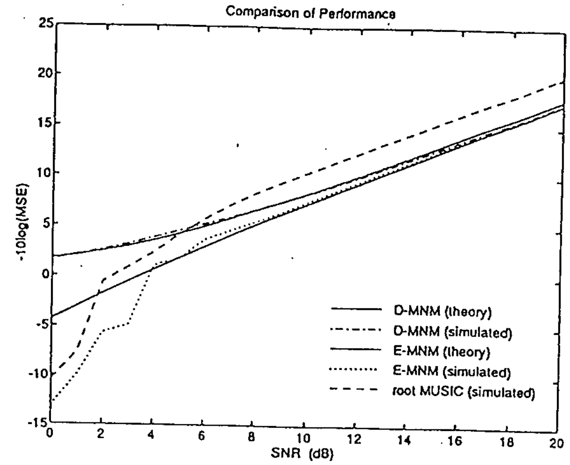


Fig.5 Performance comparision of three methods using all 200 trials.

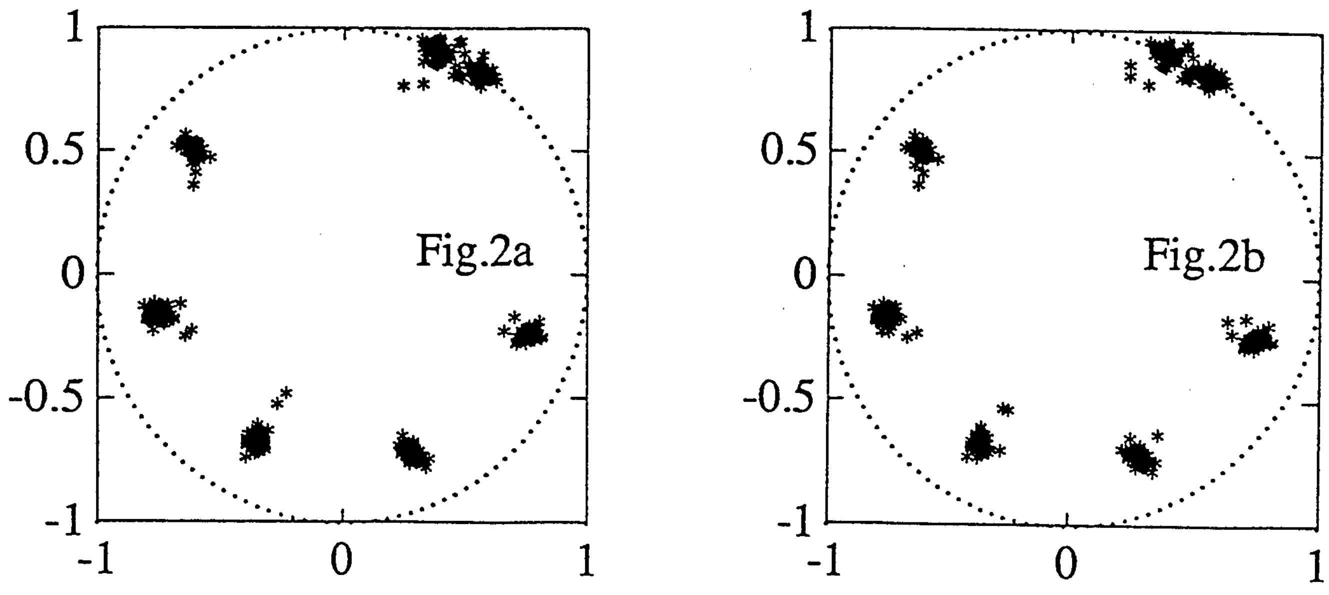


Fig.2. Roots of $D(z)$ using (a) E-MNM and (b) D-MNM for 50 independent.

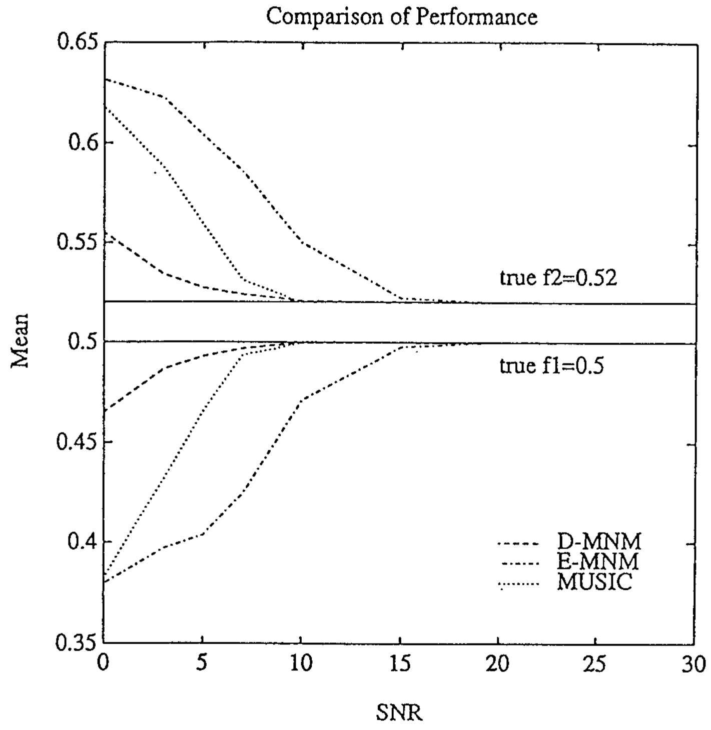


Fig.4. Comparison of Mean values with 500 independent trials for three methods.

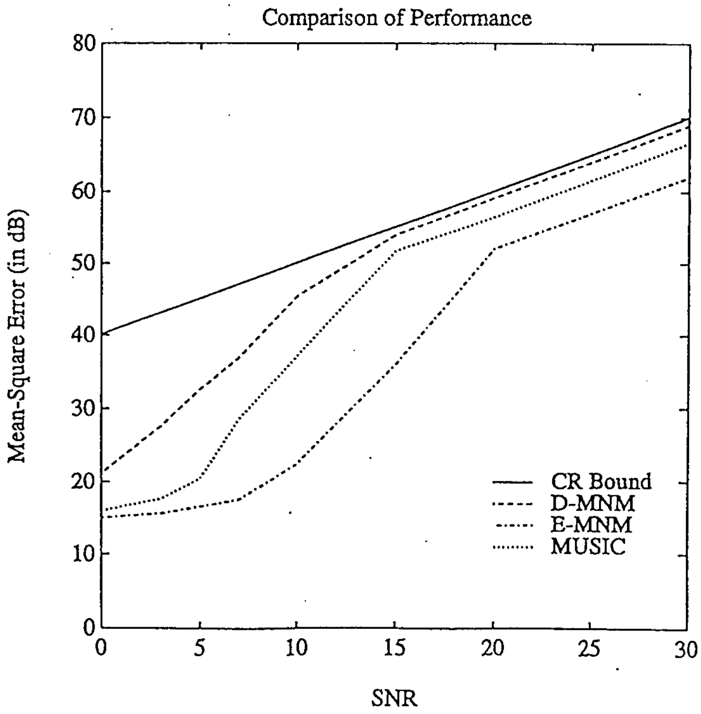


Fig.5. Comparison of RMS values with CR bounds for 500 independent trials.

SNR	Successful Trials		Bias (in degrees)		RMS	
(in dB)	D-MNM	E-MNM	D-MNM	E-MNM	D-MNM	E-MNM
5	59	39	-0.8480 1.1589	-0.5539 0.4329	1.4311 1.9174	1.3623 1.9322
10	139	130	-0.3154 0.8940	-0.4589 0.7603	1.3529 1.7910	1.5063 1.8571
15	191	189	-0.0714 0.4623	-0.1094 0.3648	0.9812 1.3118	1.0021 1.3212
20	199	198	-0.0055 0.1717	-0.0252 0.1170	0.6777 0.8440	0.6822 0.8017
25	200	200	4.99e-4 0.0611	-0.0067 0.0481	0.4129 0.4820	0.4302 0.4826
30	200	200	0.0037 0.0263	0.0018 0.0219	0.2297 0.2728	0.2329 0.2737

Table 1 : Comparison of performance of D-MNM and E-MNM.

SNR	Bias (in degrees)			RMS		
(in dB)	D-MNM	E-MNM	MUSIC	D-MNM	E-MNM	MUSIC
0	-0.0349 0.0352	-0.1205 0.1118	-0.1178 0.0983	0.0876 0.0786	0.1783 0.1748	0.1594 0.1486
3	-0.0133 0.0141	-0.1029 0.1027	-0.0681 0.0678	0.0415 0.0468	0.1654 0.1640	0.1312 0.1265
5	-0.0070 0.0072	-0.0964 0.0838	-0.0343 0.0392	0.0232 0.0342	0.1476 0.1378	0.0946 0.0991
7	-0.0031 0.0039	-0.0754 0.0658	-0.0063 0.0111	0.0142 0.0245	0.1322 0.1189	0.0373 0.0560
10	-3.40e-4 6.54e-4	-0.0289 0.0301	-5.62e-4 -1.19e-4	0.0054 0.0093	0.0756 0.0776	0.0140 0.0058
15	-7.10e-5 -1.82e-4	-0.0023 0.0019	2.80e-5 -9.05e-5	0.0020 0.0022	0.0159 0.0134	0.0026 0.0026
20	-3.40e-6 -7.76e-5	-1.04e-5 6.34e-5	1.61e-5 -5.23e-5	0.0011 0.0012	0.0025 0.0025	0.0015 0.0014
30	8.64e-6 -1.77e-5	2.18e-5 -9.01e-6	3.35e-6 -1.51e-5	3.53e-4 3.75e-4	7.87e-4 7.84e-4	4.61e-4 4.50e-4

Table 2. Comparison of Bias and RMS values for three methods with 500 independent trials.

REFERENCES

[1] G. Bienvenue and L. Kopp, "Principle de la goniometric od passive adaptive," *Proceedings 7'eme Colloque GRESTI*, Nice, France, pp. 106/1-106/10, 1979.

[2] K. Buckley and X.-L. Xu, "Spatial Spectrum Estimation in a Location Sector," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-38, no. 11, pp. 1842-1852, Nov., 1990.

[3] J. P. Burg, "Maximum Entropy Spectral Analysis," presented at the *37th Annual International SEG Meeting*, Oklahoma City, OK, 1967.

[4] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408-1418, 1969.

[5] M. P. Clark and L. L. Scharf, "Two-Dimensional Modal Analysis Based on Maximum Likelihood," To be Published, *IEEE Transactions on Signal Processing*, 1994.

[6] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Fourier Series," *Math. Comput.*, vol. 19, pp. 297-301, 1965.

[7] S. Haykin et al, Editors, *Array Signal Processing*, Prentice-Hall, 1985.

[8] Y. H. Hu, "Adaptive methods for Real Time Pisarenko Spectrum Estimate," *Proceedings of the ICASSP-1985*, March, 1985.

[9] L. B. Jackson et al., "Frequency Estimation by Linear Prediction," in the *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing-1979*, Washington, DC, pp. 352-356, Apr., 1979.

[10] D. H. Johnson et al, "Improving the Resolution of Bearing in Passive Sonar Arrays by Eigenvalue Analysis," Technical Report EE-8102, Department of Electrical Engineering, Rice University, Houston, Texas, 1980.

[11] M. Kaveh and A. J. Barbell, "The Statistical Performance of the MUSIC and the Minimum-Norm Algorithms in Resolving Plane Waves in Noise," *IEEE Transaction on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 331-341, April, 1986.

[12] A. C. Kot, S. Parthasarathy, D. W. Tufts and R. J. Vaccaro, "Statistical Performance of Single Sinusoid Frequency Estimation in White Noise Using State-Variable Balancing and Linear Prediction," to be published in *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1988.

[13] R. Kumaresan and D. W. Tufts, "A Two-Dimensional Technique for Frequency-Wavenumber Estimation," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1515-1517, Nov., 1981.

[14] R. Kumaresan and D. W. Tufts, "Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-30, no. 6, pp. 833-840, Dec., 1982.

[15] R. Kumaresan and D. W. Tufts, "Estimating the Angles of Arrival of Multiple Planewaves," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no .1, pp. 134-139, Jan., 1983.

[16] R. Kumaresan and A. K. Shaw, "High Resolution Bearing Estimation Without Eigendecomposition," *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, Florida, April, 1985.

[17] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1986.

[18] R. Kumaresan and A.K. Shaw, "Superresolution by Structured Matrix Approximation", *IEEE Transactions on Antennas and Propagation*, Vol. AP-36, pp. 34-44, 1988.

[19] S. Y. Kung, K. S. Arun and D. V. Bhaskar Rao, "State-Space and Singular-Value Decomposition-Based Approximation Methods for the Harmonic Retrieval Problem," *Journal of the Optical Society of America*, vol. 73, pp. 1799-1811, Dec., 1983.

[20] R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," *Geophysics*, vol. 36, pp. 661-675, Aug., 1971.

[21] L. C. Palmer, "Coarse Frequency Estimation Using the Discrete Fourier Transform," *IEEE Transactions on Information Theory*, Vol. IT-20, pp. 104-109, Jan., 1974.

[22] A. Paulraj, R. Roy and T. Kailath, "Estimation of Signal Parameters via Rotational Invariance Techniques - ESPRIT," *Nineteenth ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct., 1985.

[23] V. F. Pisarenko, "The Retrieval of Harmonics from Covariance Functions," *Geophysical Journal of the Royal Astronomical Society*, Vol. 33, pp. 347-366, 1973.

[24] R. Prony, "Essai Experimental et Analytique etc.," L'Polytechnique, Paris, 1 Cahier 2, pp. 24-76, 1795.

[25] B. D. Rao, "Statistical Performance Analysis of the Minimum-Norm Method," *IEE Proceedings, Part-F*, vol. 136, Pt. F, no. 3, pp. 125-134, June 1989.

[26] S. S. Reddi, "Multiple Source Location- A Digital Approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, no.1, pp. 95-105, 1979.

[27] D. C. Rife and R. R. Boorstyn, "Single Tone Parameter Estimation from Discrete-Time Observations," *IEEE Transactions on Information Theory*, Vol. IT-20, pp. 591-598, Sept., 1974.

[28] D. C. Rife and R. R. Boorstyn, "Multiple Tone Parameter Estimation from Discrete Time Observations," *Bell Systems Technical Journal*, vol. 55, pp. 1389-1410, 1976.

[29] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[30] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proceedings of RADC Spectral Estimation Workshop*, pp. 243-258, Rome, New York, 1979.

[31] A. Schuster, "On the Investigation of Hidden Periodicities with Application to a Supposed 26-Day Period of meteorological Phenomena," *Terr. Magnet.*, vol. 3, pp. 13-41, 1898.

[32] T. J. Shan , M. Wax and T. Kailath, "Spatial Smoothing Approach to Location Estimation of Coherent Sources," *Asilomar Conference on Circuits and Systems and Computers*, Pacific Grove, California, pp. 367-371, Nov., 1983.

[33] A. K. Shaw and R. Kumaresan, "Frequency-Wavenumber Estimation by Structured Matrix Approximation," *Third IEEE-ASSP Workshop on Spectrum Estimation and Modeling*, Boston, MA, pp. 81-84, Nov., 1986.

[34] A.K. Shaw, *Structured Matrix Problems in Signal Processing*, Ph.D. Dissertation, Univ. of Rhode Island, RI, 1987.

[35] A.K. Shaw and R. Kumaresan, "Some Structured Matrix Approximation Problems", *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, New York, NY, pp. 2324-2327, April, 1988.

[36] A. K. Shaw and R. Kumaresan, "Estimation of Angles of Arrivals of Broadband Sources," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, pp. 2296-2299, April, 1987.

[37] A. K. Shaw, "Approximate Maximum Likelihood Estimation of Multiple Frequencies with Constraints to

Guarantee Unit Circle Roots," *IEEE Transactions on Signal Processing,* to be published, 1994.

[38] A. K. Shaw and W. Xia, "High-Resolution Angles of Arrival Estimation using Minimum-Norm Method Without Eigendecomposition," *IEEE International Conference on Acoustics, Speech and Signal Processing,* Adelaide, Australia, April, 1994.

[39] A. K. Shaw and W. Xia, "Minimum-Norm Method Without Eigendecomposition," *IEEE Signal Processing Letters,* vol. 1, no. 1, pp. 12-14, Jan. 1994.

[40] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing,* Prentice-Hall, NJ, 1992.

[41] J. B. Y. Tsui, *Digital Microwave Receivers : Theory and Applications,* Artech House, MA, 1989.

[42] D. W. Tufts and R. Kumaresan, "Frequency Estimation of Multiple Sinusoids : Making Linear Prediction Perform Like Maximum Likelihood," *Proceedings of the IEEE,* vol. 70, pp. 975-989. Sept., 1982.

[43] D. W. Tufts and C. D. Melissinos, "Simple, Effective Computation of Principal Eigenvectors and their Eigenvalues and Application to High-Resolution Estimation of Frequencies," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-34, no. 10, pp. 1046-1053, Oct., 1986.

[44] R. J. Vaccaro, "On Adaptive Implementations of Pisarenko's Harmonic Retrieval Method," *Proceedings of the ICASSP-84,* March, 1984.

[45] H. Wang and M. Kaveh, "Estimation of Angles of Arrival for Wideband Sources," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing-1984,* San Diego, California, pp. 7.5.1-7.5.4, Mar. 19-21, 1984.

[46] H. Wang and M. Kaveh, "On the Performance of Signal Subspace Processing-Part I: Narrowband Systems," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-34, no. 5, pp. 1201-1209, Oct., 1986.

[47] X.-L. Xu and K. M. Buckley, "Bias Analysis of the MUSIC Location Estimator," *IEEE Transactions on Signal Processing,* vol. 40, no. 10, Oct., 1992.

[48] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-36, no. 10, pp. 1553-1560, Oct., 1988.

[49] M. D. Zoltowski, G. M. Kautz and S. D. Silverstein, "Beamspace Root-MUSIC," *IEEE Transactions on Signal Processing.* vol. 41, no. 1, Jan., 1993.

Section - 2.2 : Maximum-Likelihood Estimation of Multiple Frequencies with Constraints to Guarantee Unit Circle Roots

Summary

A recently proposed approximate Maximum-Likelihood Estimator (MLE) of multiple exponentials, converts the frequency estimation problem into a problem of estimating the coefficients of a z-polynomial with roots at the desired frequencies [1, 2]. Theoretically, the roots of the estimated polynomial should fall on the unit circle. But MLE, as originally proposed, does not guarantee unit circle roots. This drawback sometimes causes merged frequency estimates, especially at low SNR [1, 3]. If all the sufficient conditions for the z-polynomial to have unit circle roots are incorporated, the optimization problem becomes too nonlinear and it loses the desirable weighted-quadratic structure of MLE. In this paper, the exact constraints are imposed on each of the 1st-order factors corresponding to individual frequencies for ensuring unit circle roots. The constraints are applied during optimization *alternately* for each frequency. In the absence of any merged frequency estimates, the RMS values more closely approach the theoretical Cramer-Rao (CR) bound at low SNR levels.

I. Introduction

Estimating the underlying parameters of multiple complex exponential signals in noise remains a vigorously researched topic in signal processing [1-13]. For a single sinusoid or when the multiple frequencies are well-separated, the Periodogram performs reasonably well. But if the frequencies are closely spaced, which often occurs when the data length is limited or the aperture is too small, the Periodogram fails to distinguish the frequencies and produces merged frequency estimates. In order to overcome the Periodogram's resolution limitation, many high-resolution methods have been developed in the past two decades [1-13]. In contrast to the Periodogram, these methods make effective use of some underlying property of the true sinusoidal signal model.

Among all the existing high-resolution frequency estimation methods, the MLE appears to provide the most accurate frequency estimates and has the lowest SNR threshold [1-4]. Other high-resolution methods rely on signal or noise subspace information which is extracted from the eigendecomposition of covariance matrix or SVD of data matrix [5, 7-11]. On the other hand, the MLE considers the exact model of the exponential signal and attempts to maximize the exact likelihood function to estimate the unknowns. For a single sinusoid, the peak of the periodogram itself corresponds to the ML estimate, but for multiple exponentials the MLE turns out to be a nonlinear optimization problem [1-6, 12, 13].

The MLE approaches developed independently in [1] and [2], estimate the frequencies from the roots of a z-polynomial. It may be noted here that in literature, these methods are sometimes referred to as KiSS [1, 5, 6] or IQML [2]. In the polynomial domain, the ML optimization problem turns out to be *quasi-linear* where a weighted-quadratic criterion is minimized iteratively. Though effective to a large extent, MLE is known to possess one fundamental drawback : the optimization procedure in [1, 2] does not impose sufficient theoretical constraints on the polynomial coefficients for the estimated roots to fall on the unit circle. The primary goal of this work is to address this unresolved problem in MLE.

Two conditions must be satisfied for a general p-th order z-polynomial to have p unit circle roots : conjugate symmetry (C1) and a derivative constraint (C2), the details of which are given later. In MLE, only C1 is imposed. The derivative constraint makes the problem highly nonlinear and hence, C2 can not be incorporated in the weighted-quadratic framework of MLE. But when $p > 1$, C1 alone is not sufficient for unit circle roots. Furthermore, from the theory of Linear-Phase FIR filters, it is well-known that the roots of a symmetric z-polynomial may fall either on the unit circle or they may be in reciprocal pairs falling inside and outside of the unit circle. In fact, it was demonstrated in [1] and [3] that, if SNR $\leq$ 10dB and the frequencies are spaced closely, the roots extracted by MLE sometimes appear in reciprocal pairs. In such cases, two frequencies merge to produce only a single frequency estimate. The alternate approach proposed in this paper attempts to alleviate this limitation.

There is one exception to the two conditions stated above : for $p = 1$, the conjugate symmetry constraint (C1) alone is ***sufficient*** for the single root to fall on the unit circle. This is the main idea which will be utilized in developing the proposed Constrained-MLE (C-MLE) algorithm. Specifically, C1 will be imposed on each of the 1st-order factors of the p-th order $z-$polynomial, such that each individual root falls on the unit circle. This process need not be applied to all the frequencies at all SNRs. The constraints are imposed only on those 1st-order factors which produce merged frequency estimates at convergence of MLE. The factors for which the roots are already on the unit circle, are held fixed. The proposed algorithm may be considered to be a polynomial-domain counterpart of the 'Alternating Projection' approach [13] where the ML criterion is minimized *w.r.t.* one frequency at a time while the other frequencies are held at the previously estimated values. Our work appears to be the first attempt to guarantee unit circle roots on the polynomial coefficients for Maximum-Likelihood frequency estimation. The constraints are primarily effective at low SNR levels when there is a higher possibility for MLE to produce merged frequency estimates. In simulations, the RMS values of the frequency estimates using C-MLE were found to be closer to the theoretical CR bounds than those of the original MLE algorithm.

The paper is arranged as follows : In Section-II, the ML problem is stated and the original MLE algorithm is briefly discussed and the conditions needed for unit circle roots are stated. In Section III, the proposed constrained version of MLE is introduced. Simulation results are given in Section-IV to verify the performance of C-MLE.

II. The Maximum Likelihood Problem and a Brief Overview of MLE

The observed samples of a complex multiple exponential signal can be represented as

$$\mathbf{x}(n) \triangleq \sum_{k=1}^{p} c_k e^{j(\omega_k n \, + \, \phi_k)} + z(n) \quad n = 0,1,\ldots,N-1, \tag{1}$$

where, ω_k, c_k and ϕ_k are the unknown angular frequency, amplitude and phase, respectively, of the k^{th} sinusoid; p is the assumed number of sinusoids and $z(n)$ represents *i.i.d.* $N(0,\sigma^2)$ Gaussian noise samples. For this signal model, the MLE corresponds to optimization of the following error criterion [1-4] :

$$\min_{\omega_1,\ldots,\omega_p,a_1,\ldots,a_p} ||\mathbf{e}||^2 \triangleq \min_{\omega_1,\ldots,\omega_p,a_1,\ldots,a_p} ||\mathbf{x} - \mathbf{Ta}||_2^2 \tag{2}$$

where,

$$\mathbf{x} \triangleq \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix} \triangleq \mathbf{Ta}$$

$$\triangleq \begin{pmatrix} 1 & 1 & \ldots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \ldots & e^{j\omega_p} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\omega_1(N-1)} & e^{j\omega_2(N-1)} & \ldots & e^{j\omega_p(N-1)} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \tag{3}$$

$a_k \triangleq c_k e^{j\phi_k}$, for $k = 1,2,\ldots,p$, respectively, are the complex amplitudes. The MLE problem stated in (2) is a nonlinear optimization problem with respect to the angular frequencies. Instead, MLE forms an alternative but equivalent error criterion in the polynomial coefficient domain which has a quasi-linear structure which is well-suited for iterative optimization. A brief summary of the MLE criterion is in order.

Let, $B(z) \triangleq b_0 + b_1 z^{-1} + \ldots + b_p z^{-p}$, be a p^{th} degree z-polynomial with p roots at $e^{j\omega_1}$, $e^{j\omega_2}$ $\ldots$ $e^{j\omega_p}$, respectively, and $\mathbf{b} \triangleq [b_0 \;\; b_1 \;\; \cdots \;\; b_p]^T$ be the coefficient vector. The MLE criterion for estimating $\mathbf{b}$ is [1]-[4] :

$$\min_{\{b_i\}_{i=0}^{p}} E(\mathbf{b}) = \mathbf{b}^H \mathbf{X}^H (\mathbf{B}\mathbf{B}^H)^{-1} \mathbf{X}\mathbf{b} \quad \text{where,} \tag{4}$$

$$\mathbf{B} \triangleq \begin{pmatrix} b_p & \dots & b_1 & b_0 & & 0 \\ & \ddots & & \ddots & \ddots & \\ 0 & & b_p & \dots & b_1 & b_0 \end{pmatrix},$$

$$\mathbf{X} \triangleq \begin{pmatrix} x(p) & \dots & x(0) \\ x(p+1) & \dots & x(1) \\ \vdots & \ddots & \vdots \\ x(N-1) & \dots & x(N-p-1) \end{pmatrix}. \tag{5}$$

The criterion in (4) appears to be quadratic in $\mathbf{b}$, except that the weight matrix itself depends on the unknown coefficients. Hence, this criterion is minimized iteratively. At the $(k-1)$-th iteration

$$\min_{\mathbf{b}} \ \mathbf{b}^H[\mathbf{X}^H(\mathbf{B}^{(k-1)}\mathbf{B}^{H^{(k-1)}})^{-1}\mathbf{X}]\mathbf{b} \tag{6}$$

is optimized, where the weight matrix $(\mathbf{B}\mathbf{B}^H)$ is formed using the estimate of $\mathbf{b}$ found at the previous iteration. At convergence of these iterations, the frequencies are found from the roots of the estimated polynomial $\hat{B}(z)$. Unfortunately, direct optimization of the criterion in (4) does not guarantee that the roots of $\hat{B}(z)$ will indeed fall on the unit circle and it was recognized in [1, 3] that two conditions, must be satisfied to guarantee unit circle roots :

C1 : The coefficients possess conjugate symmetry :

$$b_k = b^*_{p-k}, \qquad \text{for, } k = 0, 1, \dots, p, \qquad \text{and,} \tag{7}$$

C2 : For $p > 1$, the derivative of $B(z)$, *i.e.*,

$$B'(z) \triangleq \frac{\partial B(z)}{\partial z^{-1}} \tag{8}$$

must have zeros either inside or on the unit circle.

The polynomial domain MLE, as originally proposed, imposes the conjugate symmetry constraint only [1, 2]. C2 makes the optimization problem highly nonlinear and the weighted-quadratic structure of (4) is lost if C2 is incorporated in the algorithm. Hence, no attempt was made in [1-4] to include C2 in the algorithm. But if $p > 1$, C1 is not a sufficient condition for unit circle roots. The same condition may, in fact, lead to roots in reciprocal pairs which can and does occur in MLE, especially at low SNR. In such cases, two closely spaced frequencies are estimated as a *single* frequency only [1, 3].

Important Observation : For $p = 1$, the conjugate symmetry alone is a *sufficient* condition to ensure unit-circle root. Hence, we propose to impose C1 sequentially on each 1st-order factor of $B(z)$ during optimization of (4). In that case, the optimization at each step will be with respect to only a 1st-order factor of $B(z)$ and hence, there is no need for satisfying C2.

III. Constrained MLE (C-MLE)

The p-th order polynomial $B(z)$ can be expressed in factored form as :

$$B(z) = B^{(p-i)}(z)B^{(i)}(z), \tag{9}$$

where, $B^{(p-i)}(z) \triangleq b_0^{(p-i)} + b_1^{(p-i)}z^{-i} + \dots + b_{p-1}^{(p-i)}z^{-p+1}$ and $B^{(i)}(z) \triangleq b_0^{(i)} + b_1^{(i)}z^{-1}$, are $(p-1)$-th order and 1st-order factors, respectively. If conjugate symmetry is imposed on the 1st order factor, then,

$B^{(i)}(z) = b_0^{(i)} + b_0^{*(i)} z^{-1}$. Note that in (9) the coefficients of the polynomial $B(z)$ are formed as the convolution of the coefficients of $B^{(p-i)}(z)$ and $B^{(i)}(z)$. Hence, in matrix-vector notation :

$$\mathbf{b} = \begin{pmatrix} b_0^{(p-i)} & 0 \\ b_1^{(p-i)} & b_0^{(p-i)} \\ \vdots & \vdots \\ b_{p-1}^{(p-i)} & b_{p-2}^{(p-i)} \\ 0 & b_{p-1}^{(p-i)} \end{pmatrix} \begin{pmatrix} b_0^{(i)} \\ b_0^{*(i)} \end{pmatrix}$$

$$\triangleq \mathbf{B}_{p-i} \begin{pmatrix} 1 & j \\ 1 & -j \end{pmatrix} \begin{pmatrix} b_{0r}^{(i)} \\ b_{0i}^{(i)} \end{pmatrix} \triangleq \mathbf{B}_{p-i}\mathbf{C}\mathbf{b}_i, \qquad (10)$$

where, $\mathbf{B}_{p-i}$ denotes the matrix-factor with the i-th 1-st order factor removed and $b_0^{(i)} \triangleq b_{0r}^{(i)} + j b_{0i}^{(i)}$. Using (10) in (6), each 1st-order factor of $B(z)$ is estimated by optimizing,

$$\min_{\mathbf{b}_i} \; \mathbf{b}_i[\mathbf{C}^H \mathbf{B}_{p-i}^{H\,(k-1)} \mathbf{X}^H (\mathbf{B}^{(k-1)} \mathbf{B}^{H^{(k-1)}})^{-1} \mathbf{X} \mathbf{B}_{p-i}^{(k-1)} \mathbf{C}] \mathbf{b}_i,$$

$$\text{for, } i = 1, 2, \ldots, p. \qquad (11)$$

This is a weighted-quadratic criterion of the form :

$$\mathbf{b}_i^H \mathbf{W}_{p-i}^{(k-1)} \mathbf{b}_i \qquad \text{where,} \qquad (12a)$$

$$\mathbf{W}_{p-i}^{(k-1)} \triangleq \mathbf{C}^H \mathbf{B}_{p-i}^{H\,(k-1)} \mathbf{X}^H (\mathbf{B}^{(k-1)} \mathbf{B}^{H^{(k-1)}})^{-1} \mathbf{X} \mathbf{B}_{p-i}^{(k-1)} \mathbf{C}$$

and

$$\mathbf{b}_i \triangleq \begin{pmatrix} b_{0r}^{(i)} \\ b_{0i}^{(i)} \end{pmatrix}. \qquad (12b)$$

Note that the weight matrix $\mathbf{W}_{p-i}^{(k-1)}$ is formed with the estimates found at $(k-1)-th$ iteration step when the unconstrained MLE algorithm is assumed to have converged. The criterion in (11) can be optimized sequentially or concurrently for each first order factor. At each iteration, $\mathbf{b}_i$ is estimated as the eigenvector corresponding to the minimum eigenvalue of $\mathbf{W}_{p-i}^{(k-1)} \in \mathbb{R}^{2\times 2}$. The advantage of using (12a) instead of (6) is that, since each $B^{(i)}(z)$ is a first-order z-polynomial, the conjugate symmetry constraint is sufficient to guarantee the root of $B^{(i)}(z)$ to fall on the unit circle. In practice, the alternate optimization procedure in (11) need not be carried out for all the p factors of $B(z)$. It needs to be invoked only in those cases for which unconstrained MLE produces merged frequency estimates. The roots which are already on the unit circle need not be optimized further. This sequential process guarantees that all the roots of $B(z)$ will indeed fall on the unit circle.

IV. Simulation Results

The algorithm described in this paper has been tested with the same simulated data set used in [1] and [2]. The following formula was used to generate the data,

$$x(n) = a_1 e^{j\omega_1 n} + a_2 e^{j\omega_2 n} + z(n) \qquad (13)$$

$$n = 0,\ 1,\ \ldots, 24$$

where, $\omega_1 = 2\pi f_1, \omega_2 = 2\pi f_2$, f_1 and f_2 being 0.52 and 0.50, respectively, $a_1 = 1, a_2 = e^{j\frac{\pi}{4}}$, $z(n)$ is a computer generated white zero-mean, complex gaussianly distributed noise sequence with variance $= \sigma^2$, *i.e.*, $\frac{\sigma^2}{2}$ is the variance of the real and the imaginary parts of $z(n)$. SNR is defined as, 10 $\log_{10}(\frac{|a_i|^2}{\sigma^2})$. Two hundred data sets with independent noise epochs were used.

Fig. 1a and 1b show the estimated roots for 200 independent trials of MLE for SNR = 5dB and 10dB, respectively. Fig. 1d and 1e show the corresponding results with C-MLE. For the 10dB case, Figures 1c and 1f show only the merged cases before after applying the exact constraints. The unit circle roots in Fig. 1f does show wider spread than the corresponding merged frequency estimates in Fig. 1c. Fig. 2 compares the performance of MLE and C-MLE with the theoretical CR bound. The results verify that C-MLE performs better than original MLE at low SNR range. The performance of C-MLE has also been compared with that of the AP method [13] and the results are displayed in Fig. 3. Clearly, the proposed method outperforms the AP method for this example, especially at low SNR.

1. R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1986.
2. Y. Bressler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 10, pp. 1081-1089, Oct., 1986.
3. A.K. Shaw, *Structured Matrix Problems in Signal Processing*, Ph.D. Dissertation, Univ. of Rhode Island, RI, 1987.
4. R. Kumaresan and A.K. Shaw, "Superresolution by Structured Matrix Approximation", *IEEE Transactions on Antennas and Propagation*, Vol. AP-36, pp. 34-44, 1988.
5. L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.
6. M. P. Clark and L. L. Scharf, "Reducing the Complexity of Parametric Estimators for Deterministic Modal Analysis," *IEEE Transactions on Signal Processing*. vol. 40, no. 7, pp. 1811-1813, July, 1992.
7. R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proceedings of RADC Spectral Estimation Workshop*, pp. 243-258, Rome, New York, 1979.
8. S. S. Reddi, "Multiple Source Location- A Digital Approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, no.1, pp. 95-105, 1979.
9. R. Kumaresan and D. W. Tufts, "Estimating the Angles of Arrival of Multiple Planewaves," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no .1, pp. 134-139, Jan., 1983.
10. S. Y. Kung, K. S. Arun and D. V. Bhaskar Rao, "State-Space and Singular-Value Decomposition-Based Approximation Methods for the Harmonic Retrieval Problem," *Journal of the Optical Society of America*, vol. 73, pp. 1799-1811, Dec., 1983.
11. A. Paulraj, R. Roy and T. Kailath, "Estimation of Signal Parameters via Rotational Invariance Techniques - ESPRIT," *International Conference on Acoustic, Speech and Signal Processing*, pp. 83-89, 1986.
12. D. C. Rife and R. R. Boorstyn, "Multiple Tone Parameter Estimation from Discrete Time Observations," *Bell Systems Technical Journal*, vol. 55, pp. 1389-1410, 1976.
13. I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-36, no. 10, pp. 1553-1560, Oct., 1988.

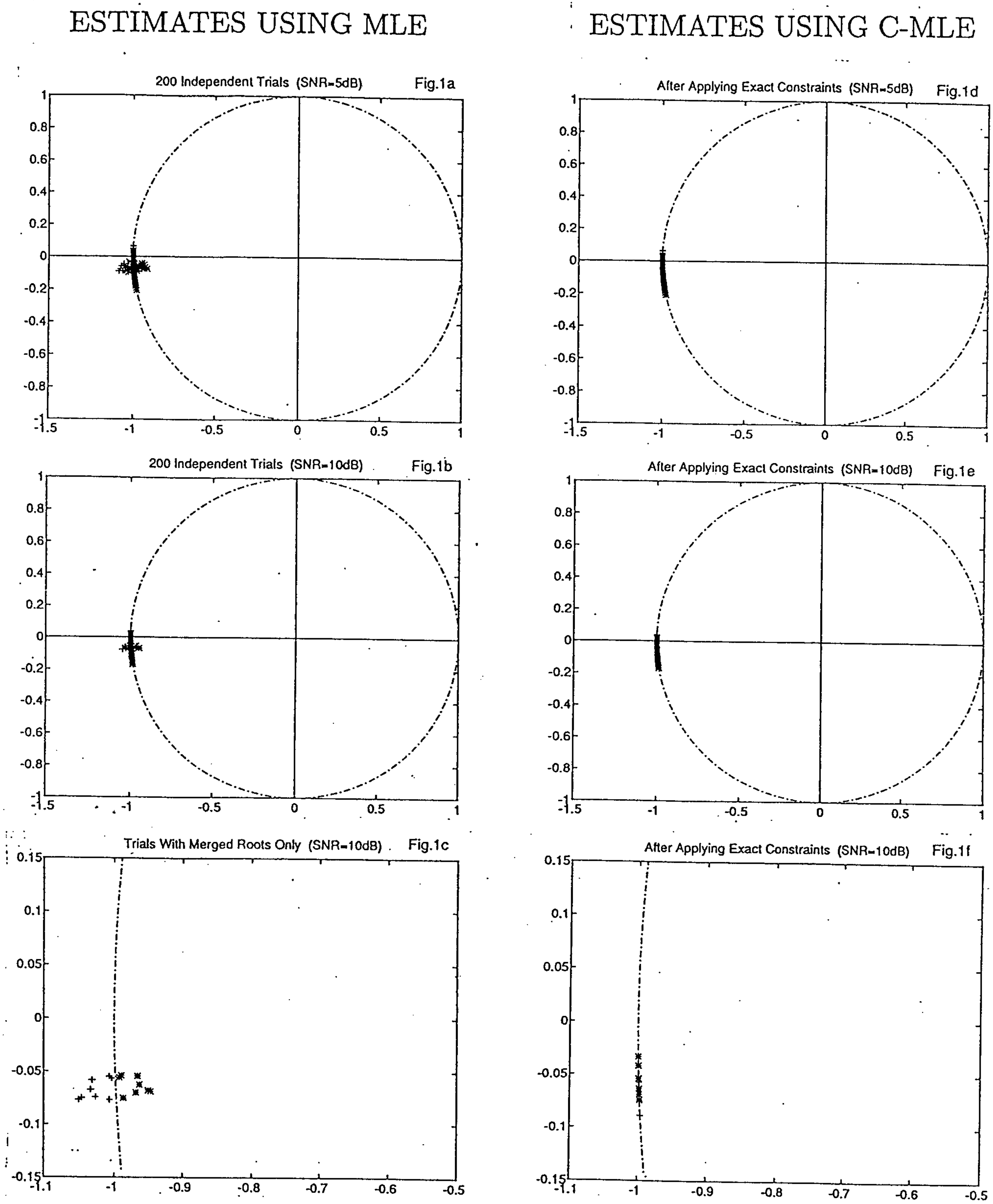


Fig. 1 : Superimposed plots of estimated roots for 200 independent trials using MLE (a-c) and C-MLE (d-f).

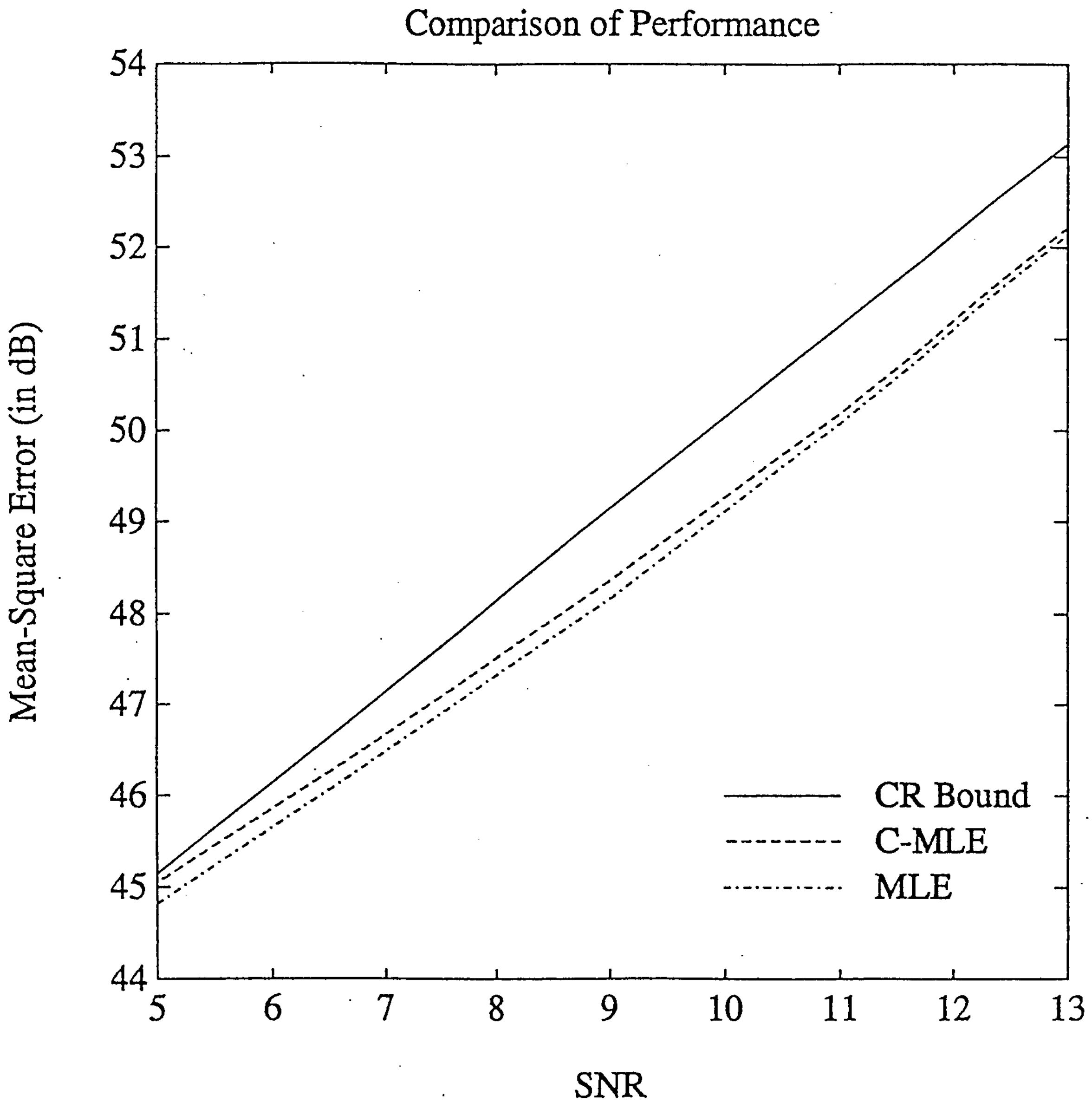


Fig. 2 : Performance comparison of MLE and C-MLE with the theoretical CR-bound. 200 independent trials were used.

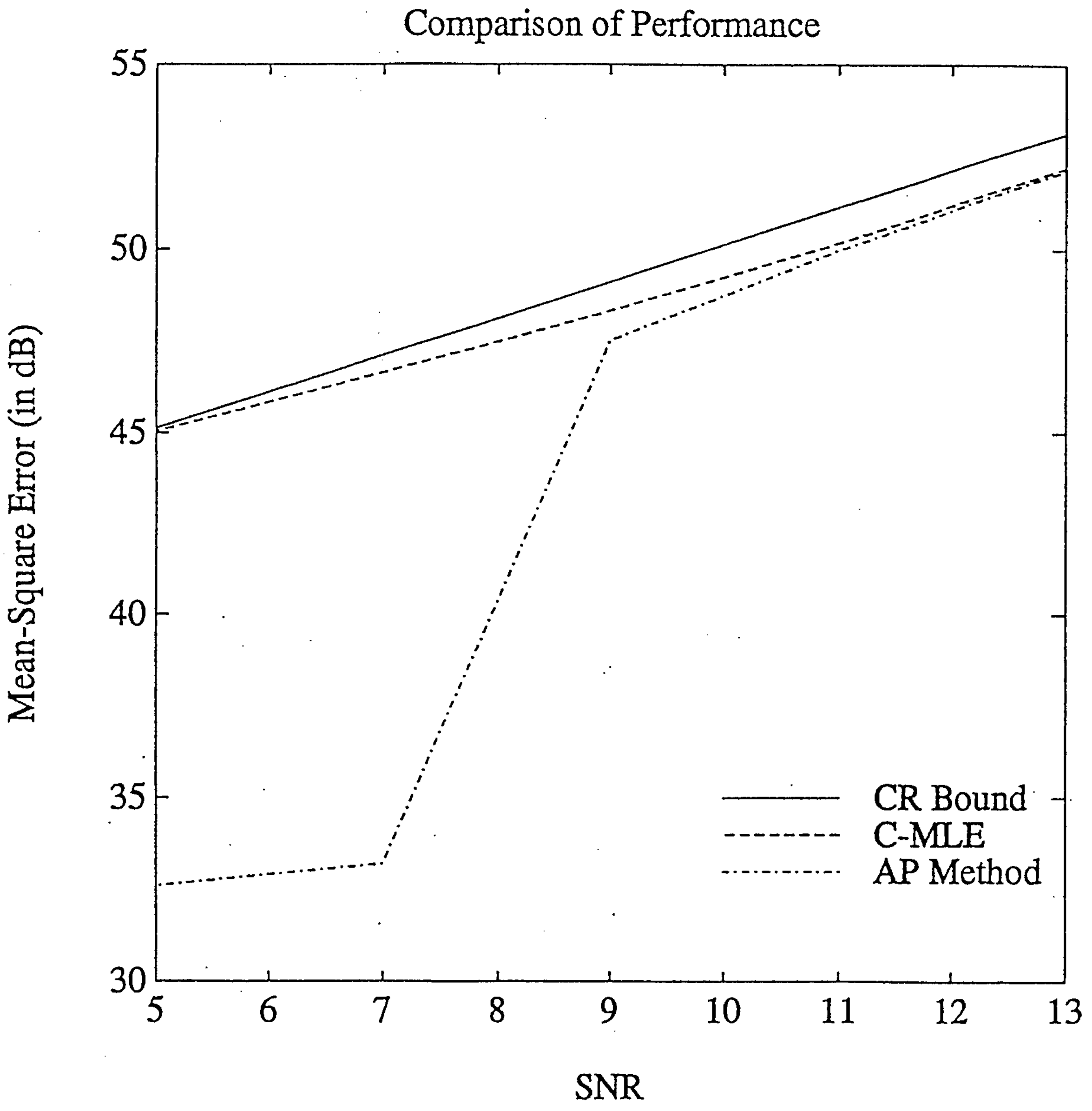


Fig. 3 : Performance comparison of C-MLE and AP methods with the theoretical CR-bound. 200 independent trials were used.

Section - 2.3 : Improved AR-Parameter Estimation from Noisy Observation Data

Summary

Auto-Regressive (AR) modeling is the most widely used approach for model-based spectrum estimation. But almost all the existing methods for AR-parameter estimation show severe degradation if the observed signal is corrupted with noise. In fact, all the commonly used techniques, such as, Autocorrelation Method (AM), Covariance Method (CM), Modified Covariance Method and their variations, give poor Power Spectral Density (PSD) estimates when the observations are noisy. In this Section, a *data-adaptive pre-filtering* approach is presented to address this problem. Preliminary results indicate that when only noisy data is available for modeling, the proposed technique gives more accurate PSD estimates than the commonly used methods.

Introduction

Auto-regressive (AR) modeling continues to play a very important role in model-based spectral estimation [1-4]. A major reason for the wide appeal of Auto-regressive (AR) modeling is its computational simplicity. Specifically, the standard AR-methods such as, Covariance method or Autocorrelation method or their variations only need to solve a set of linear equations. Furthermore, in estimating ARMA or MA models, AR-parameter estimation is a necessary intermediate step [1]. But there remains a fundamental problem with most AR-modeling methods and that is with regards to the sensitivity of the AR spectral estimators to observation noise. Noisy observation samples are indeed very common in practice, and the performance of the existing estimators deteriorate drastically in such cases. There have been some previous attempts to address this problem. AR-model in noise being a special type of ARMA model, this property has been used in [8, 9], but this makes the estimation problem highly nonlinear. Another suggested solution has been to model the process as large-order AR model so as to reduce the estimation bias, but this may lead to spurious peaks if the chosen model order is too high [11]. Other methods suggest noise compensation to remove the bias but this requires prior information about the observation noise [10].

The main goal of this Section is to utilize certain data-prefiltering ideas which have been found to be highly effective in estimating sinusoidal frequencies from noisy data [5, 6] and also for identifying deterministic systems from Input-Output data [12] and Impulse Response Data [13]. It is well-known that a sinusoidal process can be viewed as a limiting case of a narrowband AR-process. Indeed, the peak locations of AR-spectra are commonly used as the estimates of frequencies [1]. But the poor performance of AR-methods with noisy data also causes inferior frequency estimates at low SNR. In order to alleviate this problem, a large class of methods based on principal-component (PC) analysis, have been developed for reducing the effect of noise in data [3, 7]. But the PC-based methods, though highly effective for tone-frequency estimation, can not be used for cleansing noisy AR-data. This is because the data and correlation matrices are theoretically full-rank in this case even when there is no observation noise at all. A new class of algorithms, referred to as KiSS or IQML, have been developed recently for Maximum-Likelihood frequency estimation [5, 6]. The KiSS algorithm essentially prefilters the noisy data by iteratively minimizing the projection of the observations onto the noise subspace formed with linear predictor type polynomial coefficients. It is shown in this work that this matrix-prefiltering approach also has the desired noise-reduction effect on pure-AR-in-noise data. Extensive simulation studies indicate that the proposed prefiltering produces more accurate AR-spectra than the conventional AR-modeling approaches.

FORMULATION OF DATA-ADAPTIVE PREFILTERING

The proposed approach may be best explained by outlining the initialization step and the noise-subspace projection utilized by the KiSS-IQML algorithm [5, 6]. In that algorithm, the frequency estimation problem is essentially transformed into an AR-type polynomial estimation problem. Specifically, let,

$$B(z) \triangleq b_0 + b_1 z^{-1} + \ \ldots \ + b_p z^{-p} \tag{1}$$

be a p^{th} degree z-polynomial with roots at $e^{j\omega_1}$, $e^{j\omega_2}$ $\ldots$ $e^{j\omega_p}$, respectively. The coefficient vector,

$$\mathbf{b} \triangleq [b_0 \ b_1 \ \cdots \ b_p]^T \tag{2}$$

is estimated by minimization of the following error criterion :

$$\min_{\{b_i\}_{i=0}^{p}} \ \mathbf{b}^H \mathbf{X}^H (\mathbf{B}\mathbf{B}^H)^{-1} \mathbf{X}\mathbf{b} \quad \text{where,} \tag{3a}$$

$$\mathbf{B} \triangleq \begin{pmatrix} b_p & \ldots & b_0 & & 0 \\ & \ddots & & \ddots & \\ 0 & & b_p & \ldots & b_0 \end{pmatrix} \quad \text{and} \tag{3b}$$

$$\mathbf{X} \triangleq \left(\begin{array}{c|cccc} x(p) & x(p-1) & \ldots & x(0) \\ x(p+1) & x(p) & \ldots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & \ldots & x(N-p-1) \end{array}\right) \tag{3c}$$

$$\triangleq (\mathbf{g} \ \mid \ \mathbf{G}). \tag{3d}$$

The weighted-quadratic structure in (3a) is utilized for minimizing the criterion iteratively. At the $(i+1)$-th iteration, the weight matrix $(\mathbf{B}\mathbf{B}^H)$ is formed with the estimate of $\mathbf{b}$ found at the i-th iteration and the following criterion is minimized to obtain the updated estimate :

$$\min_{\mathbf{b}} \ \mathbf{b}^H [\mathbf{X}^H (\mathbf{B}^{(i)} \mathbf{B}^{H^{(i)}})^{-1} \mathbf{X}] \mathbf{b}. \tag{4}$$

The iterative algorithm in (4) is initialized with, $\mathbf{b} = [1 \ 0 \ \ldots \ 0]^T$. Hence, the initial estimator has the following form :

$$\min_{\mathbf{b}} \ \mathbf{b}^H \mathbf{X}^H \mathbf{X}\mathbf{b} \tag{5a}$$

$$= \min_{\mathbf{b}} \ ||\mathbf{X}\mathbf{b}||^2. \tag{5b}$$

Interestingly, this criterion is exactly identical to the 'Covariance Method' of linear prediction used in AR-modeling [4]. If the data contains no observation noise, the minimization in (5) would indeed produce exact frequency estimates. Furthermore, the performance of covariance method for modeling *pure* AR-processes without any observation noise is also known to be quite good [4]. But the performance deteriorates drastically with noisy observation data. In fact, simulations indicate that even at reasonably high SNR of 30-35 dB, Covariance (or Autocorrelation) method may not be able to distinguish closely spaced peaks or frequencies in the underlying process.

In order to improve on the initial estimate obtained using (5), the criterion in (4) is iteratively minimized in KiSS/IQML. But, the original criterion in (3a) also has the following equivalent forms :

$$\begin{aligned}
\mathbf{b}^H\mathbf{X}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{X}\mathbf{b} &= \mathbf{b}^H\mathbf{X}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{B}\mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{X}\mathbf{b} && (6a)\\
&= \mathbf{x}^H\mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{B}\mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{B}\mathbf{x} && (6b)\\
&= \mathbf{x}^H\mathbf{P}_{\mathbf{B}^H}\mathbf{P}_{\mathbf{B}^H}\mathbf{x} && (6c)\\
&= ||\mathbf{P}_{\mathbf{B}^H}\mathbf{x}||_2^2 && (6d)\\
&= ||\mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{B}\mathbf{x}||^2 && (6e)\\
&= ||\mathbf{W}\mathbf{B}\mathbf{x}||^2 && (6f)\\
&= ||\mathbf{W}\mathbf{X}\mathbf{b}||^2, && (6g)
\end{aligned}$$

where, $\mathbf{P}_{\mathbf{B}^H} \triangleq \mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{B}$ denotes the 'projection matrix' of $\mathbf{B}^H$,

$$\mathbf{W} \triangleq \mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1} \qquad (7a)$$

is a weighting matrix and

$$\mathbf{x} \triangleq [x(0)\ x(1)\ \ldots\ x(N-1)]^T], \qquad (7b)$$

is the observation vector.

Equation (6d) shows that, in order to reduce the effect of noise in this AR-type parameter estimates, the projection of the data ($\mathbf{x}$) onto the column-space of the $\mathbf{B}^H$ matrix needs to be minimized. The criterion in equation (6g) is similar to the criterion in (5b) for Covariance method, except that in (6g) the projection operation essentially prefilters the data matrix $\mathbf{X}$ by the weight matrix $\mathbf{W}$ which is formed by the coefficients estimated at the previous iteration step. The most obvious conclusion from this discussion is that the noise-suppression capability of KiSS-IQML is essentially due to this prefiltering of the data-matrix ($\mathbf{X}$) which appears in conventional Covariance method for AR modeling.

As mentioned before, multiple sinusoids can be modeled as a limiting case of narrowband AR-process [1]. The analogies noted above appears to lead to the possible hypothesis that similar prefiltering operation may be equally effective in reducing noise effects on the AR parameter estimates also, especially for narrowband AR-processes. The algorithm outlined next essentially minimizes the projection of the data onto the column-space of $\mathbf{B}^H$ in order to obtain improved estimates of the AR-parameter vector $\mathbf{b}$.

Steps for the Proposed Prefiltering Algorithm

1 : Obtain the initial estimate of the AR-parameters in $\mathbf{b}$ using any of the conventional AR-modeling methods.

2 : Form the $\mathbf{B}^{(i)}$ matrix defined in (3b) using the estimate of $\mathbf{b}$ found in the previous iteration.

3 : Minimize the criterion (4b) to obtain an updated estimate of $\mathbf{b}$, which has the following form :

$$\mathbf{b}^{(i+1)} = \begin{pmatrix} 1 \\ \cdots\cdots\cdots\cdots\cdots \\ -(\mathbf{G}^H(\mathbf{B}^{(i)}\mathbf{B}^{H^{(i)}})^{-1}\mathbf{G})^{-1}\mathbf{G}^H(\mathbf{B}^{(i)}\mathbf{B}^{H^{(i)}})^{-1}\mathbf{g} \end{pmatrix} \qquad (8)$$

where, $\mathbf{B}^{(i)}$ denotes the matrix obtained is Step-2 whereas, $\mathbf{g}$ and $\mathbf{G}$ are defined in (3d).

4 : Go to Step-2 unless $||\mathbf{b}^{(i+1)} - \mathbf{b}^{(i)}||^2 < \delta$, where δ is a small number.

An important difference between KiSS-IQML algorithm and the proposed method is that in case of KiSS-IQML, conjugate-symmetry constraints need to be imposed on the coefficients of the $B(z)$-polynomial in an attempt to constrain the roots to lie on the unit circle. This makes the optimization problem even more nonlinear. But in the present case no such constraints are necessary and hence, the optimization in (4) is more straightforward. Extensive simulation studies have shown that this algorithm does produce better AR spectrum match at lower SNR than any of the standard AR-modeling techniques.

SIMULATION RESULTS

The test data-set given in Chapter-7 in [3] was generated for the simulation. Fig. 1. illustrates the performance of the Covariance method for 50 independent realizations of the observation data at 30dB SNR. The solid line in Fig. 2 shows the average of estimated spectra of the 50 realizations and the dashed line shows the true spectrum. Figures 3-4 and 5-6 show the corresponding results with Modified Covariance Method and Autocorrelation Method, respectively, with identical data sets. The results clearly demonstrate that even at this moderately high SNR, none of these commonly used methods were able to distinguish the two spectral peaks for most of the noise realizations. Fig. 7 shows the results of the proposed prefiltering algorithm for those 50 identical realizations at the same SNR. The iterations converged in 6-8 iterations in all cases. Fig. 8. shows the average of the 50 realizations with the true spectrum. This improvement was found to be consistent even at lower SNR values. Similar improvements have also been observed when the Auto-correlation method and the Modified Covariance method were used to generate the initial AR parameter estimates. The plots clearly demonstrate that the proposed method was able to match the AR-spectra more closely. With simulated data, the average prediction error power for the proposed estimator was also found to be much smaller than the standard methods.

REFERENCES

[1] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[2] L.B. Jackson, *Digital Filters and Signal Processing*, Kluwer, Boston, 1986.

[3] S. M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[4] J. Makhoul, "Linear Prediction : A Tutorial Review," *Proceedings of the IEEE*, vol. 63, pp. 561-580, April, 1975.

[5] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 637-640, June 1988.

[6] Y. Bressler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 10, pp. 1081-1089, Oct. 1988.

[7] D. W. Tufts and R. Kumaresan, "Frequency Estimation of Multiple Sinusoids : Making Linear Prediction Perform Like Maximum Likelihood," *Proc. of IEEE*, vol. 70, pp. 975-989. Sept., 1982.

[8] Y. Hosoya, "Efficient Estimation of Model with an Autoregressive Signal with White Noise", *Tech. Report-37*, Dept. of Statistics, Stanford University, Mar. 1979.

[9] M. Pagano, "An Algorithm for Fitting Autoregressive Schemes", *J. Royal Stat. Soc.*, vol. 21, pp. 274-281, 1972.

[10] J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", *Proc. of IEEE*, vol. 67, pp. 1586-1604, Dec. 1979.

[11] T. J. Ulrych and R. W. Clayton, "Time series Modeling and Maximum Entropy", *Phys. Earth Planetary Int.*, vol. 12, 1976.

[12] A. K. Shaw, "A Decoupled Approach for Optimal Estimation of Transfer Function Parameters from Input-Output Data," Under 2nd review, *IEEE Transactions on Signal Processing.*

[13] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," To be published, *IEEE Transactions on Signal Processing*, Jan., 1994.

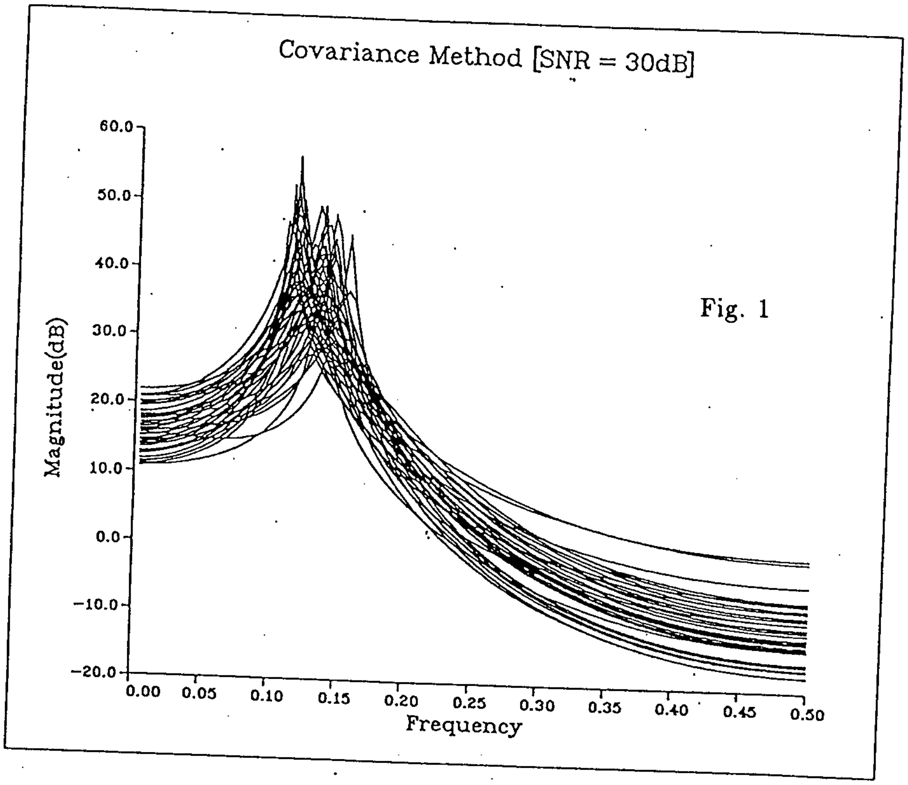

Covariance Method [SNR = 30dB]
Magnitude(dB)
60.0
50.0
40.0
30.0
20.0
10.0
0.0
−10.0
−20.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Frequency

Fig. 1

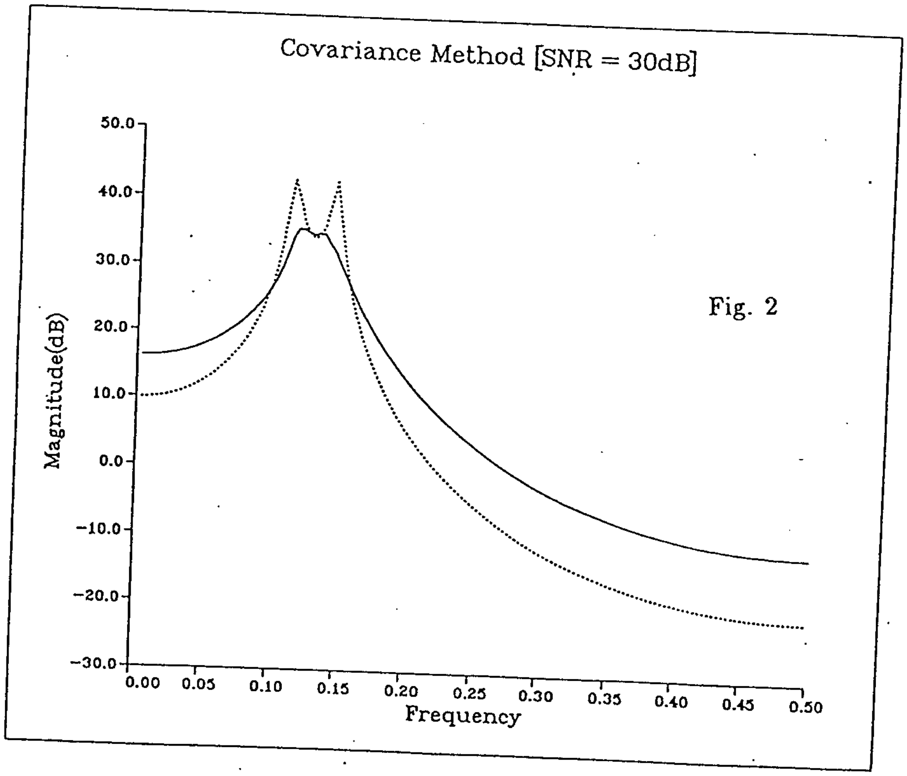

Covariance Method [SNR = 30dB]
Magnitude(dB)
50.0
40.0
30.0
20.0
10.0
0.0
−10.0
−20.0
−30.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Frequency

Fig. 2

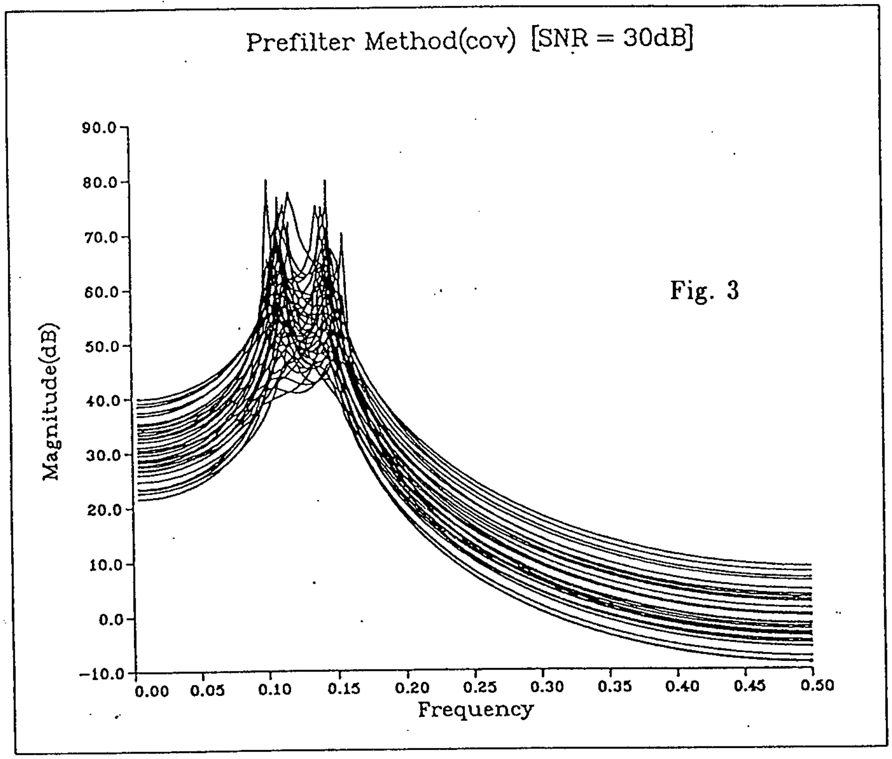

Prefilter Method(cov) [SNR = 30dB]
Magnitude(dB)
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
−10.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Frequency

Fig. 3

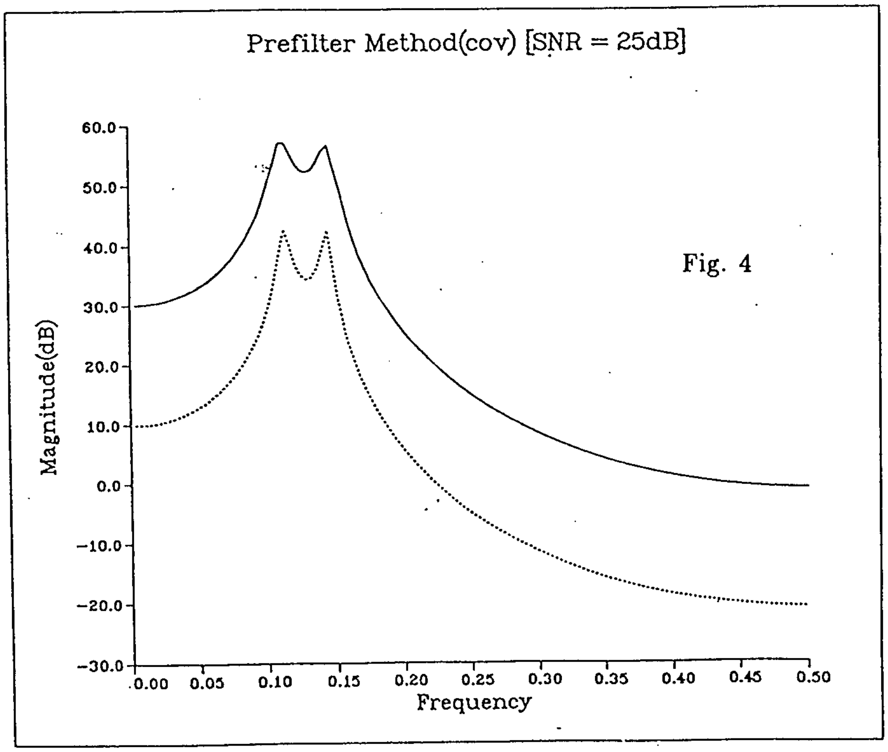

Prefilter Method(cov) [SNR = 25dB]
Magnitude(dB)
60.0
50.0
40.0
30.0
20.0
10.0
0.0
−10.0
−20.0
−30.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Frequency

Fig. 4

Section - 2.4 : Improved ARMA-Parameter Estimation from Noisy Observation Data

Summary

Existing methods for ARMA modeling assume that the available process is produced by an ARMA system driven by a white input process, *i.e.*, the observed process is considered to be pure ARMA. In practice, the available data usually have observation noise added to it but the ARMA methods do not address this problem. Simulations show that performance of the existing ARMA methods deteriorate when the observation process is noisy. In this Section a new ARMA algorithm is given which utilizes a recently developed deterministic rational system identification method (OM-IO) [8] that minimizes the modeling or output error norm. The algorithm first estimates the input process and then invokes OM-IO using the input-output data. Simulations indicate that the proposed method is quite effective even at low SNR observation data.

Introduction

With both poles and zeroes, ARMA models are best capable of effectively representing general spectra with possibly sharp peaks as well as deep valleys. Modeling of ARMA processes involves solution of a set of highly nonlinear equations. Existing methods divide the problem into several 'equation error' minimization problems to estimate the AR and MA parameters in several stages. The estimation problem is further complicated if the available data is also corrupted with observation noise. In fact, simulation studies indicate that the performance of the existing ARMA modeling methods deteriorate significantly with noisy data. This drawback may be attributed to the sensitivity of equation-error minimization based methods to the presence of noise. In this Section, we propose to address this problem by incorporating a recently developed optimal algorithm for identification of deterministic ARMA systems [8] into the stochastic ARMA modeling problem. Recent results indicate that estimators based on minimizing model 'fitting-error' have superior performance when compared to those which rely on equation-error minimization [8, 13]. In view of this, unlike existing ARMA methods, the algorithm presented in this work minimizes *output* or *modeling errors.* The results obtained so far indicate that the proposed approach is much more effective than existing methods for ARMA parameter estimation when the available data is not purely ARMA but has some observation noise added to it.

The ARMA Model

An ARMA(p,q) process can be represented in a linear difference equation form as,

$$x(n) = -\sum_{k=1}^{p} b_k x(n-k) + \sum_{k=0}^{q} a_k u(n-k) \tag{1}$$

where, the corresponding z-domain transfer function has the following form :

$$H(z) = \frac{a_0 + a_1 z^{-1} + \cdots + a_q z^{-q}}{1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_p z^{-p}} \triangleq \frac{A(z)}{B(z)}. \tag{2}$$

Let,

$$\mathbf{a} \triangleq [a_0 \; a_1 \; \cdots \; a_q]^T \qquad \text{and} \tag{3a}$$

$$\mathbf{b} \triangleq [1 \; b_1 \; \cdots b_p]^T \tag{3b}$$

denote the unknown MA and AR parameters, respectively. In vector form,

$$\mathbf{x} \triangleq [x(0) \; x(1) \cdots \; x(N-1)]^T \text{ and} \tag{4a}$$

$$\mathbf{u} \triangleq [u(0) \; u(1) \; \cdots \; u(N-1)], \tag{4b}$$

denote the output data and the driving noise sequences, respectively.

PREVIOUS METHODS

Estimation of the Maximum-Likelihood (ML) parameters of an ARMA process is a highly nonlinear problem. Akaike's MLE method requires nonlinear optimization which are prone to poor convergence if the initial estimates are chosen properly [4]. To overcome the complexities of MLE, many computationally attractive techniques have been developed also. Among these, the Modified Yule-Walker equations (MYWE) method estimates the AR parameters from the tail-end of Yule-Walker (Y-W) equations, *i.e.*, from, $r_{xx}(k), k = q+1, \ldots, q+p$. The output process is then filtered by the estimated $A(z)$ filter, which results in an MA process from which the MA parameters can be determined by any standard procedure for MA parameter estimation [9]. An extension of this approach, known as the Least-Squares MYWE (LSMYWE) [10], uses more of the tail-end of Y-W equations and yields better results than MYWE.

In stochastic ARMA modeling, the driving white noise sequence is completely unknown. Clearly, if it were somehow possible to have some estimate of the driving noise $u(n)$, then any input-output system identification technique could be used to estimate the ARMA parameters. Two well-known ARMA methods are indeed based on this principle, namely, Two-Stage Least-Squares [5] and Three-Stage Least-Squares [7, 12]. The primary steps in these methods are to model the output data first as a large order AR process, then a prediction error sequence is obtained by passing the data through the inverse filter which is MA. This whitened prediction error sequence is used as the estimate of the input white noise sequence $u(n)$. With this estimated input and the observed output, the ARMA parameters are then found by minimizing *equation errors* in two [5] or three stages [7]. The three-stage approach has been shown to have lower variance than the two-stage case. But, as will be shown with simulations below, even the three-stage algorithm can not perform well when the observation data is noisy, which is quite possible in practical situations.

It may be mentioned here that in [11] a data-adaptive prefiltering method has also been proposed for improved modeling of AR-parameters from noisy observation data. As noted in [11], there have been some previous work on AR-modeling from noisy data, but the author's are not aware of any such work for modeling ARMA parameters from noisy data, which is the problem considered in this work.

THE PROPOSED IDEA

Instead of minimizing the equation error criterion as in [5, 7], the proposed algorithm minimizes the *modeling error* or *output error* criteria. This is also a nonlinear problem, but a recently developed input-output identification method *optimally decouples* the numerator and denominator problems [8] into two separate problems of smaller dimensions. The decoupled estimators retain the global optimum of the original criterion. It has been further shown in [8] that in the decoupled form, estimation of the numerator **a** is a purely linear problem whereas the estimation of the denominator is a nonlinear problem of reduced dimensionality. But the nonlinear criterion for the denominator possesses a convenient weighted-quadratic structure which can be easily exploited to estimate the denominator iteratively. Preliminary simulation studies show that the proposed method outperforms the existing ARMA modeling approaches when the observed data is corrupted with noise. Brief explanation of the underlying theory along with the algorithm steps are in order. Some simulation results included at the end demonstrate the superior performance of the proposed method.

FORMULATION OF THE ESTIMATION PROCEDURE

Let,

$$y(n) = x(n) + v(n), \tag{5}$$

be the observed noisy ARMA process, where $v(n)$ denotes the observation noise process. Let,

$$\mathbf{y} \triangleq [y(0)\ y(1) \cdots\ y(N-1)]^T \tag{6}$$

denote the noisy output vector. Using covariance method, this observation data is first modeled as a large order (=L) AR model to obtain an AR-polynomial $B^L(z)$ such that $L >> p$, the true AR-order of the underlying ARMA process. The observation sequence $y(n)$ is then filtered through an MA-filter in the form of $B^L(z)$ to obtain a whitened prediction error sequence $\hat{u}(n)$. Then using $\hat{u}(n)$ as the estimate of the input sequence, the ARMA modeling problem can be restated as the following output-error minimization problem :

$$\min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} [y(i) - \frac{A(z)}{B(z)}\{\hat{u}(i)\}]^2. \tag{7}$$

It has been shown in [8] that the above nonlinear problem can be decoupled into a purely linear problem to estimate **a** and a nonlinear problem for **b**. Such decoupling techniques have also been found to be very effective in Maximum Likelihood estimation of the parameters of multiple exponential models [2, 3]. It may be noted here that the estimators in [6, 7] utilize the estimated prediction error sequence $\hat{u}(n)$. But those estimators do not minimize the true model-fitting defined in [7], but are based on minimizing equation error norms. The following definitions are necessary to formulate the decoupling of the numerator and denominator optimization problems.

Let $H_b(z)$ be an inverse filter corresponding to $B(z)$, *i.e.*,

$$B(z)H_b(z) = 1. \tag{8}$$

By writing this convolution in matrix-form, it can be shown that [8],

$$\mathbf{BH}_b \underline{\underline{\triangle}} \begin{bmatrix} b_{q+1} & \cdots & b_0 & 0 & \cdots & \cdots & 0 \\ \vdots & & & \ddots & \ddots & & \vdots \\ b_p & & & & b_0 & \ddots & \vdots \\ & \ddots & & & & \ddots & 0 \\ 0 & & b_p & \cdots & \cdots & \cdots & b_0 \end{bmatrix} \begin{bmatrix} h_b(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ h_b(q) & \cdots & h_b(0) \\ \vdots & \ddots & \vdots \\ h_b(N-1) & \cdots & h_b(N-q-1) \end{bmatrix} = \mathbf{0}, \tag{9}$$

which leads to the conclusion that $\mathbf{B}^T$ is orthogonal to the matrix $\mathbf{H}_b$. Utilizing this *orthogonality* relationship, the optimal criterion for estimating the denominator can be shown to be [8] :

$$\begin{aligned} &\min_{\mathbf{b}} \mathbf{y}^H\mathbf{U}_I^H\mathbf{B}(\mathbf{B}^H\mathbf{U}_I\mathbf{U}_I^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{U}_I\mathbf{y} \\ = &\min_{\mathbf{b}} \mathbf{b}^H\mathbf{Z}^H(\mathbf{B}^H\mathbf{U}_I\mathbf{U}_I^H\mathbf{B})^{-1}\mathbf{Z}\mathbf{b} \end{aligned} \tag{10}$$

where, $\mathbf{U}$ is a lower-triangular convolution matrix formed with the estimated input sequence $\hat{u}(n)$,

$$\mathbf{U} \underline{\underline{\triangle}} \begin{bmatrix} \hat{u}(0) & 0 & \ldots & 0 \\ \hat{u}(1) & \hat{u}(0) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{u}(N-1) & \hat{u}(N-2) & \ldots & \hat{u}(0) \end{bmatrix} \in \mathrm{I\!R}^{N\times N}. \tag{11a}$$

The matrix $\mathbf{U}_I$ is the inverse of $\mathbf{U}$ and is also lower triangular.

$$\mathbf{z} \underline{\underline{\triangle}} \mathbf{U}_I\mathbf{y} \quad \text{and} \tag{11b}$$

$$\mathbf{Z} \underline{\underline{\triangle}} \mathbf{B}^H\mathbf{z}, \tag{11c}$$

where, the matrix $\mathbf{Z}$ has the following Toeplitz structure,

$$\begin{aligned} \mathbf{Z} &\underline{\underline{\triangle}} \begin{bmatrix} z(q+1) & z(q) & \cdots & z(0) & 0 & \cdots & 0 \\ z(q+2) & z(q+1) & \cdots & z(1) & z(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(p) & z(p-1) & \cdots & \cdots & \cdots & \ldots & z(0) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(N-1) & z(N-2) & \cdots & \cdots & \cdots & \cdots & z(N-p-1) \end{bmatrix} \in \mathrm{I\!R}^{N-q-1\times p+1} && (11d) \\ &\underline{\underline{\triangle}} [\mathbf{g} \quad | \quad \mathbf{G}], && (11e) \end{aligned}$$

where, $\mathbf{g}$ is a column vector formed with the leading column of $\mathbf{Z}$. The denominator vector $\mathbf{b}$ is estimated by minimizing the criterion in (10). The minimization is performed iteratively by forming the weight-matrix $(\mathbf{B}^H\mathbf{U}_I\mathbf{U}_I^H\mathbf{B})$ with the estimate of $\mathbf{b}$ obtained at the previous iteration step. At convergence of the iterations, the estimated $\mathbf{b}$ is used to form the matrix $\mathbf{H}_b$ using its inverse sequence as in (8). Then the least-squares solution of the numerator $\mathbf{a}$ is found as,

$$\hat{\mathbf{a}} \triangleq (\mathbf{U}\mathbf{H}_b)^{\#}\mathbf{y} \tag{12}$$

where, $^{\#}$ denotes matrix pseudo-inverse. The iterative process is initialized by estimates obtained by minimizing equation errors as in [6, 7, 10]. Hence, the further iterations of the proposed method can only improve upon the equation-error based estimates because it minimizes the true modeling error criterion defined in (7).

The Overall algorithm in Brief

The complete algorithm for the ARMA parameter identification can be summarized as the following four primary steps :

1. Model the observed sequence $\mathbf{y}$ by a large order AR model.

2. Determine the prediction error white noise sequence $\hat{\mathbf{u}}$, which is treated as the input sequence for the system.

3. Knowing $\hat{\mathbf{u}}$ and $\mathbf{y}$, start the iterative procedure to minimize the error criterion in (10). At each iteration, $\mathbf{b}$ is estimated either as the eigenvector corresponding to the minimum eigenvalue of $\mathbf{Z}^H(\mathbf{B}^H\mathbf{U}_I\mathbf{U}_I^H\mathbf{B})^{-1}\mathbf{Z}$ or by setting $b_0 = 1$. In the later case, the estimate of $\mathbf{b}$ at the $(i+1)$-th iteration is obtained using the estimates of the previous iteration step as follows,

$$\mathbf{b}^{(i+1)} = \begin{bmatrix} 1 \\ \cdots\ \cdots\ \cdots\ \cdots \\ -(\mathbf{G}^H\mathbf{W}^{H^{(i)}}\mathbf{W}^{(i)}\mathbf{G})^{-1}\mathbf{G}^H\mathbf{W}H^{(i)}\mathbf{W}^{(i)}\mathbf{g} \end{bmatrix}, \tag{13a}$$

   where, the matrix $\mathbf{W}$ is formed with the estimates of $\mathbf{b}$ at the previous iteration step as,

$$\mathbf{W} \triangleq \mathbf{U}_I^H(\mathbf{B}^H\mathbf{U}_I\mathbf{U}_I^H\mathbf{B})^{-1}. \tag{13b}$$

   The iterations are continued till convergence is reached, *i.e.*, no significant change is found in $\mathbf{b}$ between successive iterations.

4. Estimate $\mathbf{a}$ using (12).

SIMULATION RESULTS

The simulations were performed with the test data set (ARMA4) used in Chapter-10 of [1]. The true system PSD has two prominent peaks as shown in Fig. 1. The PSD estimates at an SNR of 20dB in the observation data $y(n)$ are shown in Fig. 2. The results using MYWE, LSMYWE, Maine-Firoozan and the proposed method are shown in Figures 2a through 2d. Clearly the proposed method performs better than the other three. The corresponding results with 15dB SNR are shown in Fig. 3a - 3d. The performance of the proposed method is maintained even at this level of SNR, though the results with the three existing methods have deteriorated. Further simulations at lower SNR levels indicate that the peaky spectral shape is maintained at least up to 12dB. The efficacy of the algorithm is obvious.

REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[2] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 637-640, June 1988.

[3] Y. Bressler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 10, pp. 1081-1089, Oct. 1988.

[4] H. Akaike, "Maximum Likelihood Identification Gaussian Autoregressive Moving Average Models," *Biometrika*, vol. 60, pp. 255-265, 1973.

[5] J. Durbin, "The Fitting of Time-Series Models," *J. Roy. Stat. Soc.*, vol. 22, pp 233-243, 1960

[6] T. C. Hsia, *System Identification*, D. C. Heath, Lexington, MA, 1977.

[7] D. Q. Mayne and F. Firoozan,"An Efficient, Multistage, Linear Identification Method for ARMA Processes," *Proceeding of IEEE Conference on Decision and Control*, New Orleans, Vol. 1, pp. 435-438, Dec. 1977.

[8] A. K. Shaw, "A New Algorithm for Optimal Estimation of Plant Parameters from Input-Output Data," *31st IEEE Conference on Decision and Control*, Tucson, AZ, pp. 1684-1685, Dec., 1992.

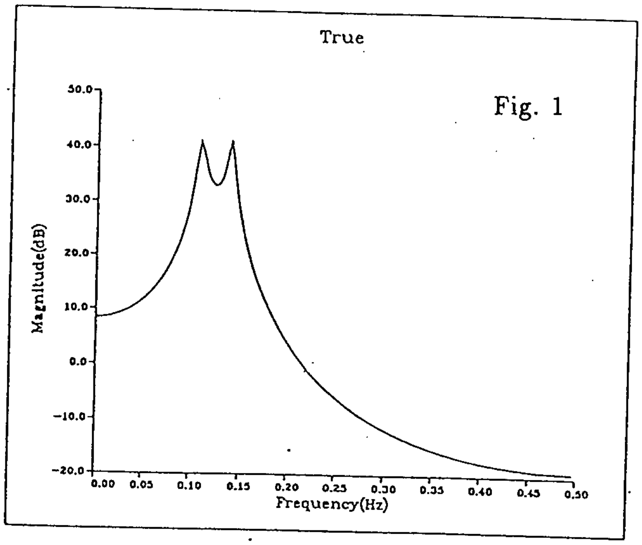
[9] J. Durbin, "Efficient Estimation of Parameters in Moving Average Models," *Biometrika* vol. 46, pp 306-316, 1959.

[10] J. Cadzow, "High Performance Spectral Spectral Estimation - A New ARMA Method," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-28, pp 524-529, Oct. 1980.

[11] A. K. Shaw and S. Kundu, "AR-Spectrum Estimation from Noisy Observation Data," Submitted for Review, *27th Asilomar Conference on Signals, Systems, and Computers.*

[12] I. S. Konvalinka and M. R. Matausek, "Simultaneous Estimation of Poles and Zeros in Speech Analysis and ITIF-Iterative Inverse Filtering Algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-27, pp 485-492, Oct. 1979.

[13] A. K. Shaw, "Optimal Estimation of the Parameters of All-Pole Transfer Functions," *IEEE Transactions on Circuits and Systems*, vol. 41, no. 2, pp.



True

Fig. 1

50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-20.0
Magnitude(dB)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Frequency(Hz)

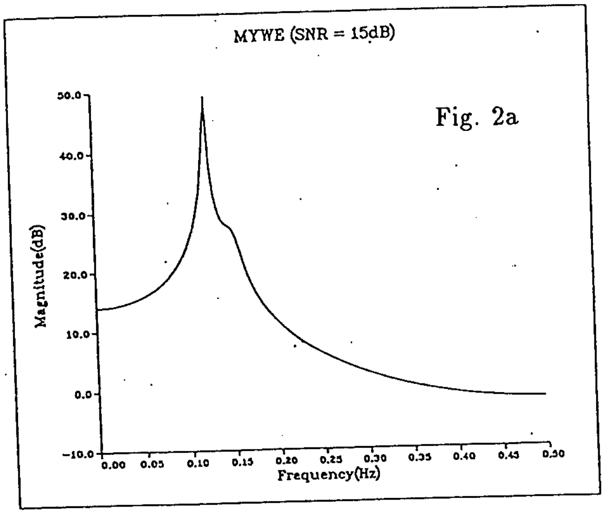


Fig. 2a

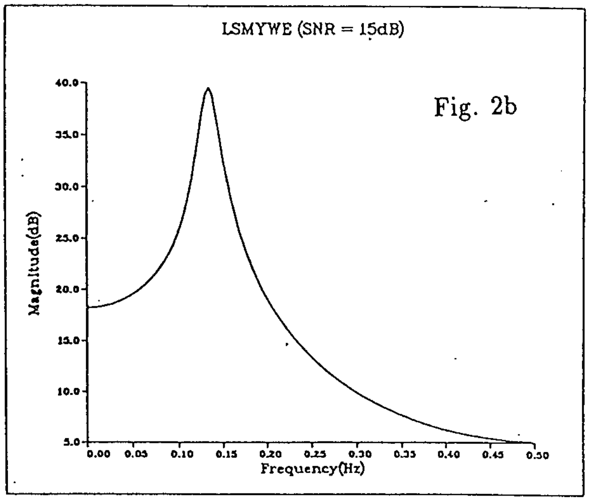


Fig. 2b

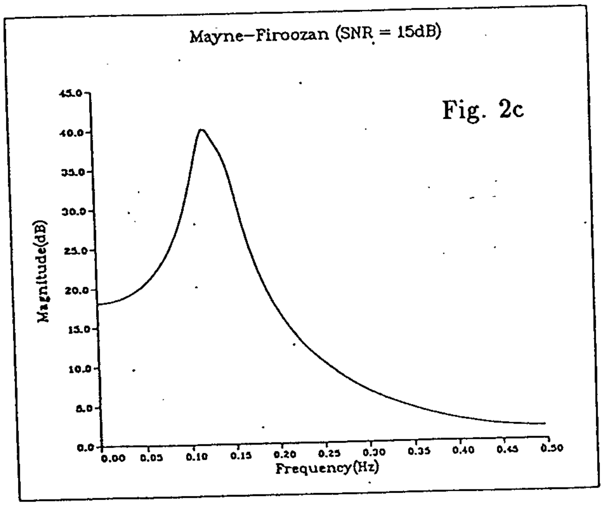


Fig. 2c

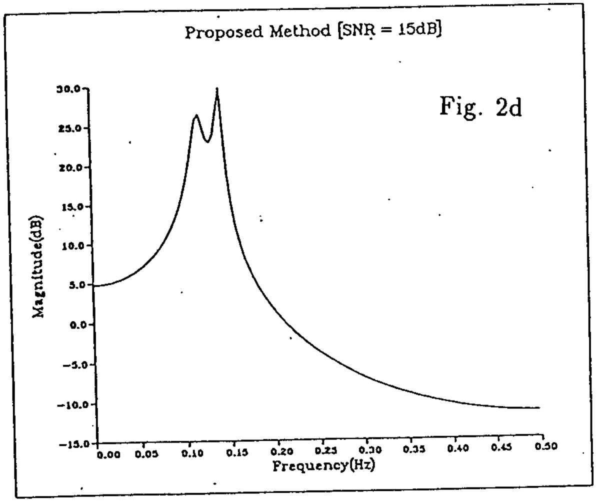
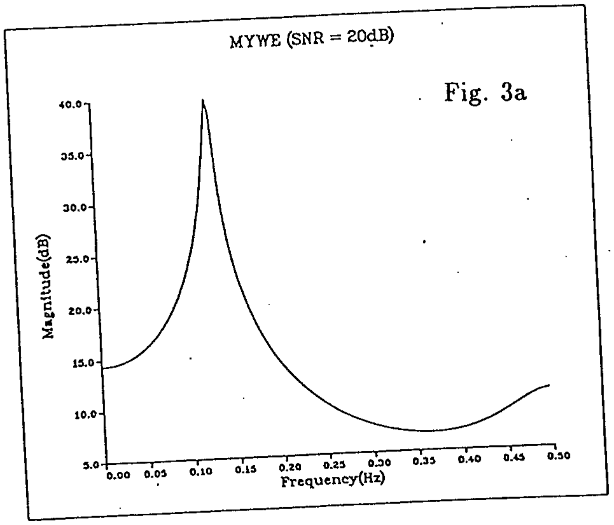
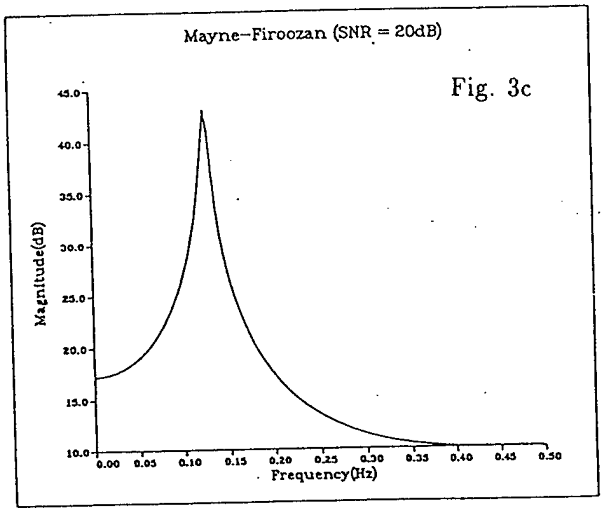


Fig. 2d



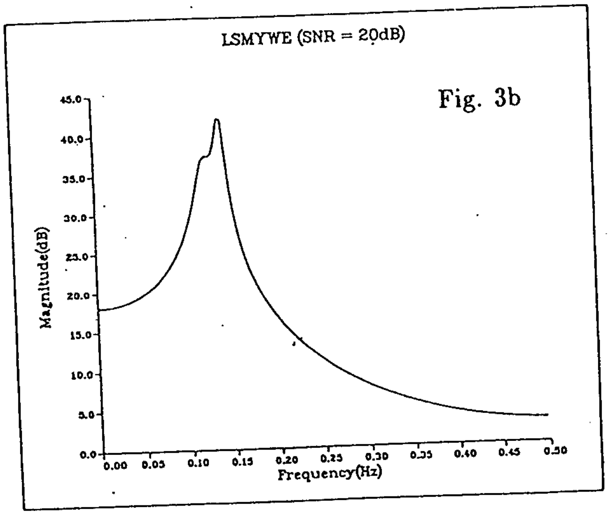
MYWE (SNR = 20dB)

Fig. 3a



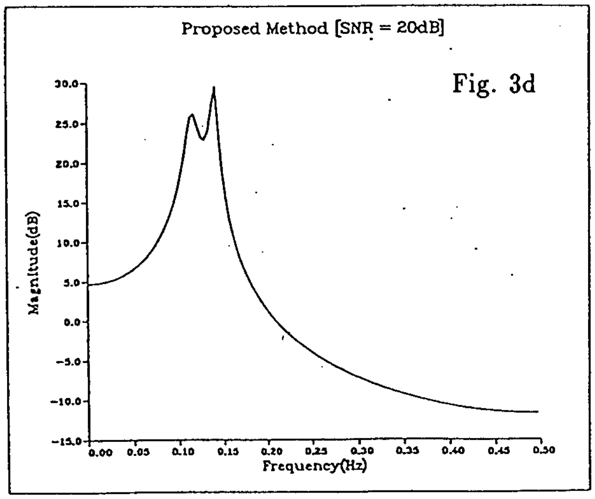
Mayne–Firoozan (SNR = 20dB)

Fig. 3c



LSMYWE (SNR = 20dB)

Fig. 3b



Proposed Method [SNR = 20dB]

Fig. 3d

Section - 2.5 : Time-Domain Detection of Electronic Warfare Signals in Noise

Summary

In the passive mode of operations of EW applications, source signals may not be present within a given observation window, or the signals may fill only a part of the estimation window. In either case, any frequency estimation algorithm may produce erroneous or noise frequencies. Considering the relatively high computational burden of any frequency estimation method, it is desirable to invoke a frequency estimation method only when a detection scheme indicates high probability of presence of threat. In this Section, the theory of detection of sinusoids from Quantized and Noisy time-domain observation samples have been developed. The theoretical work on single/multiple samples is mostly complete. Studies with Quantized data have also been performed and the results appear reasonably good. Lab tests for the Envelope Detection and Square-Law cases have been conducted at Wright Labs with satisfactory results.

I. Introduction :

In Electronic Warfare (EW) environments, microwave receivers play a major role in *passive* identification and localization of unknown targets emitting high-frequency electro-magnetic signals. EW signals cover a relatively wide bandwidth, typically in the range of 0.2 to 15 GHz, and existing microwave receivers utilize mostly analog signal processing tools and techniques [1-3]. In fact, there are no EW receivers that process microwave radar signals entirely in the digital domain. With the emergence of increasingly faster and inexpensive digital computers and high-speed A/D converters, it is expected that digital processing of microwave radar signals is expected to be practically feasible.

The primary task of a microwave receiver is to gather data for sorting of signals and identification of the type of the radar emitting the received signal. Based on these information, jamming, weapon delivery or other options are considered. In order to perform these tasks, the receiver must analyze the received radar pulses and measure or estimate the following six parameters : Angle-of-Arrival (AOA), Radio Frequency (RF), Time of Arrival (TOA), Pulse Amplitude (PA), Pulse Width (PW) and Polarization (P). But in order to reduce computational burden, the estimation of these parameters should be undertaken only when it is determined that there is a high probability of the presence of a threat signal.

In this part of the project, the detection problem has been considered in the time-domain for single and multiple samples. Detection thresholds and Probability of Detection based on Neyman-Pearson Criterion have been derived. Derivations are given for calculating the Thresholds and Probability of Detection for both the 'Square-Law' and 'Envelope' detectors.

II. Time-Domain Detection :

Almost all existing AOA/RF estimation algorithms assume that the signal is already present in the observed data. But in the passive mode of operations of EW applications, source signals may not be present at all within the observation window, or the signals may fill only a part of the estimation window. In either case, any frequency estimation algorithm would essentially produce erroneous or noise frequencies because the observed signal would not satisfy the model assumed by the estimation algorithm. Considering the relatively high computational burden, any estimation method should be invoked only when a detection scheme indicates high probability of presence of threat.

Since EW receivers do not have any prior knowledge about the frequency/amplitude/phase of the received signals, conventional *matched filters* can not be used in this case. An obvious solution would be to perform the detection in the frequency-domain, *i.e.*, the presence of targets can be determined by thresholding of FFT-peaks. The frequency-domain approaches are robust but have certain disadvantages in that a decision can be made only after a block of data has been collected. Furthermore, a lot of computational power may be wasted if FFT is taken continuously, even when no target is present. Instead, we plan to incorporate a time-domain detection scheme that can detect targets in real-time using a single observation or a small number of samples. Once a preliminary decision is taken, FFT or more sophisticated frequency/AOA estimation algorithm can be invoked, if desired.

II.1 : Signal and Noise Model

Microwave radars signals can be modeled as,

$$\begin{aligned} x(n) &= A\cos(\omega_c k + \theta) + n(k) && (1a)\\ &= A\cos(\omega_c k)\cos\theta - A\sin(\omega_c k)\sin\theta + n_I(k)\cos(\omega_c k) + n_Q(k)\sin(\omega_c k) && (1b)\end{aligned}$$

where, $n(k)$ denotes narrowband noise samples. To perform the time-domain detection, the received real data is first converted into a complex analytic signal. This is achieved by passing the real signal through a Hilbert Transformer to form the in-phase (I) and quadrature (Q) components of the complex analytic signal. When no signal is present, the I and Q components may be represented as,

$$\begin{aligned} X_I(k) &= n_I(k) && (2a)\\ X_Q(k) &= n_Q(k). && (2b)\end{aligned}$$

On the other hand, in the presence of signal, the corresponding components are given as :

$$\begin{aligned} X_I(k) &= A\cos\theta + n_I(k) && (3a)\\ X_Q(k) &= A\sin\theta + n_Q(k). && (3b)\end{aligned}$$

Since the amplitude, frequency and the phase of the received signal are unknown, the detection criterion has to rely on thresholding of the amplitude (PA) of the analytic signal. The frequency and phase can be ignored for detecting only the presence of a target signal. The amplitude threshold can not be based on minimizing the total probability of error because the exact amplitude of the signal is unknown at the receiver. Furthermore, the probability of False Alarm (P_{FA}) must also be kept very low (10^{-6} or smaller). Hence, the best detection scheme would be to calculate the threshold by setting the P_{FA} to a constant. The thresholds for Square-Law detector have been derived next for single and multiple samples within a pulse.

II.2 : Square Law Detector

The noise is assumed to be narrowband and Gaussianly distributed with zero-mean and variance $= \sigma^2$. Hence, for the no-signal case of (2) the I/Q noise samples are distributed as :

$$\begin{aligned} X_I(k) &= N(0,\sigma^2) && (4a)\\ X_Q(k) &= N(0,\sigma^2). && (4b)\end{aligned}$$

In the following derivation, the time-variable k will be suppressed until the multiple samples case is considered.

II.2.a : Single Sample Case

Assuming independent noise samples, when no signal is present, the joint probability density function (PDF) of the I/Q channel outputs are given by :

$$f(X_I, X_Q) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x_I^2 + x_Q^2)}. \qquad (5)$$

Let,

$$\begin{aligned} X_I &= R\cos\alpha && (6a)\\ X_Q &= R\sin\alpha. && (6b)\end{aligned}$$

Using the Jacobian of this transformation, the joint PDF for this polar form can be shown to be :

$$f(r,\alpha) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r). \tag{7}$$

From this the marginal for the Envelope (R) is given by,

$$f_R(r) = \int_0^{2\pi} f(r,\alpha) d\alpha = \frac{r}{2\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r), \tag{8}$$

which is known as the Rayleigh PDF.

II.2.a.1 : The PDF and Characteristic Function with No Signal

A square-law detector forms the following quantity,

$$Z \underset{=}{\triangle} X_I^2 + X_Q^2 = R^2 \tag{9}$$

which needs to be compared to a threshold to decide the presence/absence of a radar target. Since, $\frac{dZ}{dR} = 2R = 2\sqrt{Z}$, the PDF of the Square-Law output when no signal (denoted as, $\bar{s}$) is given as :

$$f_Z(z|\bar{s}) = \frac{f_R(\sqrt{z})}{2\sqrt{z}} = \frac{1}{2\sigma^2} e^{-\frac{z}{2\sigma^2}} u(z), \tag{10}$$

which is the Exponential PDF. The Characteristic Function (CF) is defined as the Fourier Transform of the Density function :

$$\begin{aligned} \mathbf{C}_Z^{\bar{s}}(\omega) &\underset{=}{\triangle} \mathcal{F}[f_Z(z|\bar{s})] \\ &= \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{z}{2\sigma^2}} \; e^{-j\omega z} dz \\ &= \frac{1}{1 + j2\omega\sigma^2} \end{aligned} \tag{11}$$

II.2.a.2 : The PDF and Characteristic Function in Presence of Signal

When target is present, *i.e.*, in case of (3), the I/Q samples are distributed as :

$$X_I(k) = N(A\cos\theta, \sigma^2) \tag{12a}$$

$$X_Q(k) = N(A\cos\theta, \sigma^2). \tag{12b}$$

In this case, the joint probability density function (PDF) is given by :

$$f(x_I, x_Q) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(x_I - A\cos\theta)^2 + (x_Q - A\sin\theta)^2]}. \tag{13}$$

Once again, using the Jacobian of the transformation, the polar-form joint PDF can be shown to be :

$$f(r,\alpha|s) = \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[r^2 + A^2 - 2Ar\cos(\alpha - \theta)]} u(r). \tag{14}$$

Integrating over α, the marginal PDF of the Envelope is given by,

$$f_R(r|s) = \int_0^{2\pi} f(r,\alpha|s) d\alpha \tag{15a}$$

$$= \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(A^2 + r^2)} \int_0^{2\pi} e^{-\frac{2Ar}{2\sigma^2}\cos(\alpha - \theta)} d\alpha, \tag{15b}$$

$$= \frac{r}{\sigma^2} e^{-\frac{1}{2\sigma^2}(A^2 + r^2)} I_0\left(\frac{Ar}{\sigma^2}\right) \tag{15c}$$

where, $I_0(\cdot)$ denotes Bessel Function of the zero-th kind. The PDF in (15c) is known as the Rician.

Similar to the no-signal case in (9)-(10), the PDF of the Square-Law output Z with signal-plus-noise is given as :

$$f_Z(z|s) \;=\; \frac{f_R(\sqrt{z})}{2\sqrt{z}} \;=\; \frac{1}{2\sigma^2} e^{-\frac{(A^2 \,+\, z)}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right). \tag{16}$$

In this case, the Characteristic Function can be found as follows :

$$\begin{aligned} \mathbf{C}_Z^s(\omega) \;&\triangleq\; \mathcal{F}[f_Z(z|s)] \\ &= \frac{1}{2\sigma^2} e^{-\frac{A^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{z}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right) e^{-j\omega z} dz \end{aligned} \tag{17}$$

The following Fourier Transform pair can be found in [CF, page-79, pair 655.1] :

$$\mathcal{F}\left[e^{-\rho g} I_0\left(\frac{2\sqrt{g}}{\sqrt{\lambda}}\right)\right] \;\longleftrightarrow\; \frac{1}{\omega \,+\, \rho} e^{\frac{1}{\lambda(\omega \,+\, \rho)}}. \tag{18}$$

Using (18) and with appropriate substitutions, $C_Z^s(\omega)$ is given by :

$$\mathbf{C}_Z^s(\omega) \;=\; \frac{1}{1 \,+\, j2\omega\sigma^2} e^{-\frac{j\omega A^2}{1 \,+\, j2\omega\sigma^2}} \tag{19}$$

II.2.a.3 : The Neyman-Pearson Criterion with a Single Sample :

For this one-dimensional case, the decision that the signal is present is taken if the likelihood-ratio [17] :

$$\ell \;=\; \frac{f_Z(z|s)}{f_Z(z|\overline{s})} > k(P_{FA}) \tag{20}$$

where k is a constant that depends on the probability of False-Alarm P_{FA}. From this relationship it may appear that in order to find the decision threshold, one would need to know or estimate the signal. But one of the most attractive consequence of Neyman-Pearson criterion is that for a given predetermined P_{FA}, the threshold can be set by integrating $f(z|\overline{s})$ over the region where the signal is present [11, 17].

II.2.a.4 : Probability of False Alarm

If the threshold is denoted as γ, the false-alarm probability can be calculated as,

$$\begin{aligned} P_{FA}^1 \;&=\; \int_\gamma^\infty f_Z(z|\overline{s})dz \\ &= \frac{1}{2\sigma^2}\int_\gamma^\infty e^{-\frac{z}{2\sigma^2}} dz \qquad \text{from (10),} \\ &= e^{-\frac{\gamma}{2\sigma^2}} \end{aligned} \tag{21}$$

II.2.a.5 : Detection Threshold

Taking natural logarithm of both sides of (21), the detection threshold is given as,

$$\gamma \;=\; -\,2\sigma^2 \ln P_{FA}^1. \tag{22}$$

II.2.a.6 : Probability of Detection

If the square-law output z of is greater than γ from (22), then the decision is taken that source target is present. Hence the probability of detection can be calculated from :

$$\begin{aligned} P_D^1 &= \int_\gamma^\infty f_Z(z|s)dz \\ &= 1 - \int_0^\gamma f_Z(z|s)dz \\ &= 1 - \frac{1}{2\sigma^2} e^{-\frac{A^2}{2\sigma^2}} \int_0^\gamma e^{-\frac{z}{2\sigma^2}} I_0\left(\frac{A\sqrt{z}}{\sigma^2}\right) dz. \end{aligned} \tag{23}$$

By letting, $v^2 = \frac{z}{2\sigma^2}$ and with appropriate substitutions,

$$P_D^1 = 1 - e^{-\frac{A^2}{2\sigma^2}} \int_0^{\sqrt{\frac{\gamma}{2\sigma^2}}} v e^{-v^2} I_0\left(2\sqrt{\frac{A^2}{2\sigma^2}}v\right) dv. \tag{24}$$

But this integral possesses the form of an Incomplete Toronto Function [13, pp-348] which is defined as follows :

$$T_B(m,n,r) \triangleq 2r^{n-m+1}e^{-r^2} \int_0^B t^{m-n} e^{-t^2} I_n(2rt)dt \tag{25}$$

Hence P_D^1 can be written in a more compact form as :

$$P_D^1 = 1 - T_{\sqrt{\frac{\gamma}{2\sigma^2}}}\left(1, 0, \sqrt{\frac{A^2}{2\sigma^2}}\right). \tag{26}$$

II.2.b : Multiple Samples Case

The detector performance can be expected to improve if the decisions can be based on multiple observations within a pulse. The question would then be how to combine the multiple samples in order to come up with an inference. For the Envelope Detection case, Tsui and Sharpin have recently derived an M-out-of-N scheme where the presence of target is decided if at least M samples out of a total of N exceed the detection threshold [12]. In this work we take a different approach where decisions are taken based on the sum of N squared samples. This approach is more akin to traditional CW detection schemes where integration over N pulses is performed for making a decision [13].

Let Y be the random variable formed with the sum of N independent squared samples, *i.e.*,

$$Y \triangleq \sum_{k=1}^{N} Z(k), \tag{27}$$

where, the PDF and CF of $Z(k)$ were derived in II.2.a.

II.2.b.1 : The PDF and Characteristic Function of Y with No Signal

When no signal is present, the PDF of Y which is formed as the sum of N *independent* samples, is given by the following convolution :

$$f_Y(y|\bar{s}) \triangleq f_Z(z_1|\bar{s}) \star f_Z(z_2|\bar{s}) \star \dots \star f_Z(z_N|\bar{s}) \tag{28}$$

where, each of the $Z(k)$'s has identical distribution. Direct convolution of N PDFs appears complicated, but it is well-known that convolution in PDF-domain implies multiplication in the CF-domain. Consequently, the CF of Y is given by,

$$\begin{aligned} C_Y^{\bar{s}}(\omega) &= \prod_{k=1}^{N} C_Z^{\bar{s}}(\omega) = [C_Z^{\bar{s}}(\omega)]^N \\ &= \frac{1}{(1 + j2\omega\sigma^2)^N} \qquad \text{using (11)} \end{aligned} \tag{29}$$

Using the inverse Fourier Transform pair-431 [Campbell and Foster, pp-44], the PDF of Y is,

$$f_Y(y|\overline{s}) \;=\; \frac{y^{N-1}e^{-\frac{y}{2\sigma^2}}}{(2\sigma^2)^N(N-1)!}u(y) \tag{30}$$

II.2.b.2 : The PDF and Characteristic Function of Y in Presence of Signal

Using arguments similar to those in the previous subsection, when signal is present, the CF of Y is given by,

$$\begin{aligned} C_Y^s(\omega) \;&=\; \prod_{k=1}^{N} C_Z^s(\omega) \;=\; [C_Z^s(\omega)]^N \\ &=\; \frac{1}{(1\;+\;j2\omega\sigma^2)^N}e^{-\frac{jN\omega A^2}{1\;+\;j2\omega\sigma^2}} \end{aligned} \tag{31}$$

Once again, using the inverse Fourier Transform pair-650.0 [Campbell and Foster, pp-77], the PDF of Y is,

$$f_Y(y|s) \;=\; \frac{1}{2\sigma^2}\frac{y}{NA^2}^{\frac{N-1}{2}}e^{-\frac{1}{2\sigma^2}(NA^2\;+\;y)}I_{N-1}\left(\frac{A\sqrt{Ny}}{\sigma^2}\right)u(y) \tag{32}$$

II.2.b.3 : The Neyman-Pearson Criterion with Multiple Samples :

For this N-dimensional case, the decision that the signal is present is taken if the likelihood-ratio [17] :

$$\ell_Y \;=\; \frac{f_Y(y|s)}{f_Y(y|\overline{s})} > k(P_{FA}). \tag{33}$$

For a given predetermined P_{FA}, the threshold can be set by integrating $f(y|\overline{s})$ over the region where the signal is present [11, 17].

II.2.b.4 : Probability of False Alarm

For γ denoting the threshold, the false-alarm probability is,

$$\begin{aligned} P_{FA}^N \;&=\; \int_\gamma^\infty f_Y(y|\overline{s})dy \\ &=\; \frac{1}{2\sigma^2}\int_\gamma^\infty \frac{e^{-\frac{y}{2\sigma^2}}y^{N-1}}{(N-1)!}dy \qquad \text{from (30)}, \\ &=\; 1\;-\;\frac{1}{2\sigma^2}\int_0^\gamma \frac{e^{-\frac{y}{2\sigma^2}}y^{N-1}}{(N-1)!}dy && (34a) \\ &=\; 1\;-\;I\left(\frac{d}{2\sigma^2\sqrt{N}}, N-1\right) && (34b) \end{aligned}$$

where, $I(\cdot)$ denotes Incomplete Gamma Function which is defined as,

$$I(u,t) \;\triangleq\; \int_0^{u\sqrt{1+t}} \frac{e^{-v}v^t}{t!}dv. \tag{35}$$

II.2.b.5 : Detection Threshold

For a given P_{FA}, the threshold γ can be determined numerically with a computer or using available plots/tables [13].

II.2.b.6 : Probability of Detection

If the sum-of-squares y is greater than the threshold γ determined from (35), then the decision is taken that source target is present. Hence the probability of detection can be calculated from :

$$\begin{aligned} P_D^N &= \int_\gamma^\infty f_Y(y|s)dy \\ &= 1 - \int_0^\gamma f_Y(y|s)dy \\ &= 1 - \frac{e^{-\frac{NA^2}{2\sigma^2}}}{2\sigma^2}\left(\frac{2\sigma^2}{NA^2}\right)^{\frac{N-1}{2}} \int_0^\gamma {\frac{y}{\sigma^2}}^{\frac{N-1}{2}} e^{-\frac{y}{2\sigma^2}} I_{N-1}\left(\frac{A\sqrt{Ny}}{\sigma^2}\right)dy \end{aligned} \tag{36}$$

By letting, $v^2 = \frac{y}{2\sigma^2}$ and with appropriate substitutions,

$$P_D^N = = 1 - \frac{e^{-\frac{NA^2}{2\sigma^2}}}{2\sigma^2}\left(\frac{2\sigma^2}{NA^2}\right)^{\frac{N-1}{2}} \int_0^{\sqrt{\frac{\gamma}{2\sigma^2}}} v^{N-1} e^{-v^2} I_{N-1}\left(2\sqrt{\frac{NA^2}{2\sigma^2}}v\right)dv. \tag{37}$$

This integral also possesses the form of an Incomplete Toronto Function defined in (25). Hence P_D^N can be written in a more compact form as :

$$P_D^N = 1 - T_{\sqrt{\frac{\gamma}{2\sigma^2}}}\left(2N-1, N-1, \sqrt{\frac{NA^2}{2\sigma^2}}\right). \tag{38}$$

II.3 : Envelope Detector

The PDF of the envelope (R) for a single sample was found in (8). Hence, for a given P_{FA}, the detection threshold is,

$$\gamma = \sqrt{-2\sigma^2 \ln P_{FA}}. \tag{39}$$

The calculation of threshold with N observation samples can be shown to be [12],

$$\gamma = T\sqrt{N\left(2 - \frac{\pi}{2}\right)} + N\sqrt{\frac{\pi}{2}} \quad \text{where,} \tag{40a}$$

T is found approximately from,

$$P_{FA} = 0.5(1 - \phi^{-1}(T)) \tag{40b}$$

and $\phi(\cdot)$ denotes the error function. More details for the Envelope Detector case can be found in [12].

It may be noted that unlike the square law and envelope detection threshold calculations for conventional radars [13], the discretized schemes presented here do not use matched filtered output but use the sampled data directly.

References

[1] J. B. Y. Tsui, *Microwave receivers with Related Components*, National Technical Information Center, 1982, Peninsula, Los Altos, CA, 1985.

[2] J. B. Y. Tsui, *Microwave Receivers with Electronic Warfare Applications*, John Wiley and Sons., New York, 1986.

[3] D. Curtis Schleher, *Introduction to Electronic Warfare*, Artech House, MA, 1986.

[4] J. B. Y. Tsui, *Digital Microwave Receivers : Theory and Applications*, Artech House, MA, 1989.

[5] J. Y. Cheung, "A Direct Adaptive Frequency Estimation Technique," *30th Midwest Symposium on Circuits and Systems,* New York, Aug., 1987.

[6] S.M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[7] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol.ASSP-34, pp. 637-640, June, 1988.

[8] Y. Bressler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-34, no. 10, pp. 1081-1089, Oct., 1988.

[9] A. K. Shaw, "A Novel Cyclic Algorithm for Maximum-Likelihood Frequency Estimation," *IEEE International Conference on Systems Engineering*, Dayton, OH, Aug., 1991.

[10] D. W. Tufts and R. Kumaresan "Frequency Estimation of Multiple Sinusoids : Making Linear Prediction Perform Like Maximum Likelihood," *Proceedings of the IEEE*, vol. 70, pp. 975-989. Sept., 1982.

[11] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[12] J. B. Y. Tsui and D. Sharpin, Unpublished Report on Time-Domain Detection for Digital Receivers, June, 1992.

[13] J. V. DiFranco and W. L. Rubin, *Radar Detection,* Artech House, Inc., Dedham, MA.

[14] J. I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar," *Trans. IRE Prof. Group on Information Theory*, IT-6; vol. 2, pp. 59-267, April, 1960.

[15] G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications,* Van Nostrand, Princeton, NJ, 1948.

[16] W. C. Knight, R. G. Pridham and S. M. Kay, "Digital Signal Processing of Sonar," *Proceedings of the IEEE,* vol. 69, no. 11, Nov. 1981.

[17] M. Schwartz and L. Shaw, *Signal Processing : Discrete Spectral Analysis, Detection, and Estimation,* McGraw-Hill, New York, 1975.

Section - 2.6 : PIPELINED-ADAPTIVE TRACKING OF MULTIPLE SINUSOIDAL FREQUENCIES

SUMMARY

New Pipelined-Adaptive algorithms are proposed for tracking multiple Frequencies or Angles-of-Arrival (AOA) of moving targets. Pipelining of adaptive filters pose a critical challenge because of the timing mismatch arising from the feedback signals. In this paper, some relaxation techniques [9] will be utilized to pipeline adaptive algorithms for high-speed tracking of frequency/AOAs. Two adaptive tracking algorithms have been mapped into pipelined forms, namely Least-Mean Squares (LMS) [3] and Recursive Least-Squares (RLS). Preliminary simulation studies with multiple sources indicate encouraging results.

I. Introduction : Pipelined data-adaptive algorithms are presented for passive high-speed tracking of multiple targets. In non-stationary environment or when target locations change with time, block-mode processing of observation data is inappropriate while adaptive algorithms are more preferable. Various adaptive algorithms addressing this problem exist [3,5] but the throughput rate of these algorithms are limited by usually long critical paths of the adaptive filters. Critical paths can be reduced by pipelining which is usually accomplished by introducing appropriate latches at intermediate stages to divide the critical path into multiple disjoint sections. Pipelining allows higher sampling rate and throughput essential in many radar applications such as in digital microwave receivers [11]. However, pipelining of adaptive filters pose additional challenge due to timing mismatch produced by feedback signals [9]. Recently, some relaxation techniques have been found to be effective in pipelining certain adaptive algorithms for coding and communication applications [9]. In this work, we study the effectiveness of relaxations for pipelining adaptive tracking algorithms.

It may be noted that various adaptive algorithms for tracking multiple targets do exist, including LMS [3], gradient adaptive lattice (GAL) [3], least squares lattice (LSL) [3], Recursive Least Squares (RLS) and QR-based adaptive tracking algorithms [5]. However, to the best of our knowledge, none of these adaptive frequency tracking algorithms have been implemented or studied in pipelined forms. Here we present the results of relaxation-based pipelining on LMS and RLS based tracking algorithms. Research on pipelining of the other tracking algorithms is being conducted and will be reported later. It may be emphasized here that pipelining will not only be beneficial for speeding up adaptive tracking, recent studies indicate that pipelining can be also effective for reduction of both power consumption [1] and chip-area [8] using appropriate folding techniques.

II. Look-Ahead Pipelining (LAP) : Consider the first order recursive equation given by,

$$y(n+1) = ay(n) + x(n). \tag{1}$$

The corresponding transfer function is given by,

$$H(z) = \frac{z^{-1}}{1 - az^{-1}} \tag{1b}$$

By applying an M-step look-ahead using back-substitution,

$$y(n) = a^M y(n-M) + \sum_{i=0}^{M-1} a^i x(n-i-1). \tag{2}$$

It can be easily shown that the transfer function corresponding to both (1) and (2) are identical. Note that $y(n)$ no longer depends on the previous output sample $y(n-1)$ but on an output that is M samples back in time, *i.e.*, $y(n-M)$. Hence, the immediate dependence problem has been removed [4, 6], *i.e.*, the signals can be sampled more often or there is more time for computation. This implies that the throughput of the logic unit is increased by a factor of M leading to high-speed implementation. This technique is referred to as time-domain [6, 10] or Clustered *look-ahead* pipelining [7] Note also that speed-up in throughput is achieved at the cost of higher hardware complexity due to the overhead term in (2).

III. Relaxation Techniques : The above LAP scheme maintains exact equivalence in the transfer function. However, the second term in (2) is an overhead term and from the standpoint of hardware, exact computation of this overhead term may be impractical, particularly for real-time adaptive implementations. It has been shown in [9] for a variety of coding and communication related applications, that these overhead terms can be approximated under certain circumstances.

III.1. Sum Relaxation : In (2), if $a \approx 1$, and if $x(n)$ remains approximately constant over M clock cycles, then we can replace the overhead term in $y(n+M)$ by [9],

$$y(n+M) = a^M y(n) + Mx(n) \tag{3}$$

III.2. Product Relaxation : When a is time-varying (represented more appropriately by $a(n)$), but its magnitude is close to unity, then $a(n)$ can be written as $(1 - a'(n))$ where $a'(n)$ is close to zero. The equation for $y(n+M)$ can be approximated as [9],

$$y(n+M) = (1 - Ma'(n))y(n) + \sum_{i=0}^{M-1} a^i(n)x(n+M-1-i). \tag{4}$$

IV. Adaptive Frequency Tracking using Pipelined LMS Adaptive Filters : Consider the un-pipelined LMS algorithm which is referred to as the *'serial'* LMS (SLMS) algorithm

$$\hat{E}_o^e(n) = (1 - \alpha_{\mathrm{LMS}})\hat{E}_o^e(n-1) + \alpha_{\mathrm{LMS}} u^2(n) \tag{5}$$

$$\mu(n) = \frac{\alpha_{\mathrm{LMS}}}{p\hat{E}_o^e(n)} \tag{6}$$

$$\mathbf{W}(n) = \mathbf{W}(n-1) + \mu(n)e(n)\mathbf{U}(n) \tag{7}$$

$$e(n) = d(n) - \mathbf{W}^T(n-1)\mathbf{U}(n) \tag{8}$$

where, $0 < \alpha_{\mathrm{LMS}} < 1$ has been used, p is the order of the transversal filter and $\hat{E}_o^e(n)$ is the power of the signal samples within the tracking window [3]. By applying M-step look-ahead to equations (5) and (7), we have

$$\hat{E}_o^e(n) = (1 - \alpha_{\mathrm{LMS}})^M \hat{E}_o^e(n-M)$$

$$\alpha_{\mathrm{LMS}} \sum_{i=0}^{M-1} (1 - \alpha_{\mathrm{LMS}})^i u^2(n-i).(9)$$

$$\mathbf{W}(n) = \mathbf{W}(n-M) + \mu(n) \sum_{i=0}^{M-1} e(n-i)\mathbf{U}(n-i). \tag{10}$$

This introduces M latches in the recursive loops which may be redistributed to pipeline the feedback multiply-add operation by M levels. By substituting the equation (10) into equation (8) we obtain the error equation

$$e(n) = d(n) - \left[\mu(n) \sum_{i=0}^{M-1} e(n-i-1)\mathbf{U}(n-i-1)\right]^T \mathbf{U}(n) - \mathbf{W}^T(n-M-1)\mathbf{U}(n). \tag{11}$$

Clearly, the number of overhead terms after applying look-ahead is rather high. By applying the sum relaxation to equation (11) and replacing $\mathbf{W}(n-M-1)$ by $\mathbf{W}(n-M)$ we can approximate equation (11) as,

$$e(n) = d(n) - \mathbf{W}^T(n-M)\mathbf{U}(n). \tag{12}$$

The sum relaxation is applied assuming that $\mu(n)$ is relatively small and hence the second term in equation (11) does not have a dominating effect on the error calculation. We also approximate equation (9) to

$$\hat{E}_o^e(n) = (1 - \alpha_{\text{LMS}})^M \left[\hat{E}_o^e(n-M) + \alpha_{\text{LMS}} \sum_{i=0}^{M-1} u^2(n-i) \right] . \tag{13}$$

Equations (6), (10), (12) and (13) constitute the relaxed look-ahead pipelined LMS algorithm (PLMS). It may be noted here that the PLMS version given in [9] does not make use of or pipeline this adaptive error calculation given in (13) which has been shown to be convenient in adaptive tracking [3].

Note that the hardware complexity after relaxed look-ahead pipelining has increased by $(N+1)(M-1)$ adders because of the overhead terms in (9) and (10). The architectures of the SLMS and PLMS filters are as shown Fig. 1a and Fig. 1b, respectively. By comparing the critical paths of the two architectures we see that by proper distribution of the extra delays introduced by pipelining, the pipelined architecture can be made to operate approximately M times faster.

V. Adaptive Frequency Tracking using Pipelined Recursive Least-Squares Algorithm : The 'serial' recursive least-squares (SRLS) algorithm is described by [2,9] :

$$\begin{aligned}
\mathbf{k}(n) &= \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{u}(n)}{1 + \lambda^{-1}\mathbf{u}^T(n)\mathbf{P}(n-1)\mathbf{u}(n)} \\
\alpha(n) &= d(n) \; - \; \mathbf{W}^T(n-1)\mathbf{u}(n); \\
\mathbf{W}(n) &= \mathbf{W}(n-1) \; + \; \mathbf{k}(n)\alpha(n); \\
\mathbf{P}(n) &= \lambda^{-1} \left[\mathbf{P}(n-1) - \mathbf{k}(n)\mathbf{u}^T(n)\mathbf{P}(n-1) \right] ;
\end{aligned}$$

where, $\mathbf{u^T(n)} = [u(n), u(n-1), \cdots, u(n-N+1)]$ is the input vector, $\mathbf{W^T(n)} = [w_1(n), \cdots, w_N(n)]$ is the vector of weights, $d(n)$ is the desired signal, $\mathbf{k}(n)$ is the Kalman filter gain, $\alpha(n)$ is the error and $\mathbf{P}(n) = \phi^{-1}(n)$, $\phi(n)$ being the deterministic autocorrelation matrix of the input signal. $\phi(n) \underline{\underline{\triangle}} \sum_{i=0}^{n} = \lambda^{n-i}\mathbf{u}(i)\mathbf{u}^T(i)$. This RLS algorithm solves for $\mathbf{W}(n)$ in the following normal equations,

$$\phi(n)\mathbf{W}(n) = \theta(n)$$

where, the cross-correlation vector $\theta(n) \underline{\underline{\triangle}} \sum_{i=0}^{n} \lambda^{n-i} d(i)\mathbf{u}(i)$. $\theta(n)$ and $\phi(n)$ can be computed recursively as,

$$\begin{aligned}
\theta(n) &= \lambda\theta(n-1) \; + \; d(n)\mathbf{u}(n); \\
\phi(n) &= \lambda\phi(n-1) \; + \; \mathbf{u}(n)\mathbf{u}^T(n)
\end{aligned}$$

Relaxed look-ahead pipelining is applied to the above two recursive equations to obtain

$$\begin{aligned}
\theta(n) &= \lambda\theta(n-M) \; + \; L_A d(n)\mathbf{u}(n); \\
\phi(n) &= \lambda\phi(n-M) \; + \; L_A \mathbf{u}(n)\mathbf{u}^T(n);
\end{aligned}$$

where, L_A is the look-ahead factor. The sum and product relaxations were used to obtain the pipelined equations. Using these equations to solve for $\mathbf{W}(n)$, the pipelined RLS (PRLS) equations can be re-derived [9] and are given as,

$$\begin{aligned}
\mathbf{k}(n) &= \frac{\lambda^{-1}\mathbf{P}(n-M)\mathbf{u}(n)}{1 + \lambda^{-1}\mathbf{u}^T(n)\mathbf{P}(n-M)\mathbf{u}(n)} \\
\alpha(n) &= d(n) \; - \; \mathbf{W}^T(n-M)\mathbf{u}(n) \\
\mathbf{W}(n) &= \mathbf{W}(n-M) \; + \; \mathbf{k}(n)\alpha(n) \\
\mathbf{P}(n) &= \lambda^{-1} \left[\mathbf{P}(n-M) - \mathbf{k}(n)\mathbf{u}^T(n)\mathbf{P}(n-M) \right]
\end{aligned}$$

The architecture for both SRLS and PRLS are given in Fig. 2a and Fig. 2b, respectively. We can see that the hardware complexity is almost the same as that of the serial algorithm except for an additional $2M$ latches. One set of M latches corresponds to that required to pipeline the W-loop and the other to pipeline the P-loop. The M latches can be redistributed within the architecture so as to maximize throughput. Furthermore, by employing the folding transformation [8], the hardware of the pipelined algorithm can be further reduced.

VI. Simulation Results : Several simulations have been conducted to verify the performance of the pipelined adaptive algorithm in tracking time varying frequencies at various noise levels.

Simulation 1 : The data set consists of 2 real sinusoids of 0.4375 Hz and 0.1250 Hz which undergo a step change to 0.3750 Hz and 0.0625 Hz respectively. The signals are at signal to noise ratios (SNRs) of 20 dB and 15 dB respectively. Fig. 3a and 3b show the tracking characteristics of the SLMS (when M = 1) and PLMS (with M = 3), respectively, keeping the new parameter at $\alpha = 0.04$. Fig. 4a and 4b show the corresponding results using SRLS (with M = 1) and PRLS (with M = 3), respectively, with $\lambda = 0.95$. Clearly, in both cases convergence for the target with higher SNR is quicker than the low SNR target. Furthermore, there is little effect on convergence time due to relaxations used in pipelining especially in the case of the PLMS. In case of PRLS, there appears to be some jitter when trying to keep up with the change for the lower SNR target.

Simulation 2 : The effectiveness of PLMS and PRLS in tracking time-varying frequencies has also been tested by letting the algorithm track two sinusoidal FM signals ($f_{c1} = 0.3750Hz$, $f_{c2} = 0.1250Hz$ and $\Delta f_1 = \Delta f_2 = 0.0625Hz$), where f_c and Δf represent center frequency and peak frequency deviation of an FM signal, respectively. Both signals are at 20 dB. The simulations are shown in Fig. 5a and 5b for the SLMS case (M = 1) and the PLMS case (M = 3) respectively with $\alpha = 0.15$. Fig. 6a and 6b show the simulation results for the SRLS (M =1) and PRLS (M=3) cases respectively with $\lambda = 0.7$. Again the PLMS and the PRLS show minimal convergence degradation.

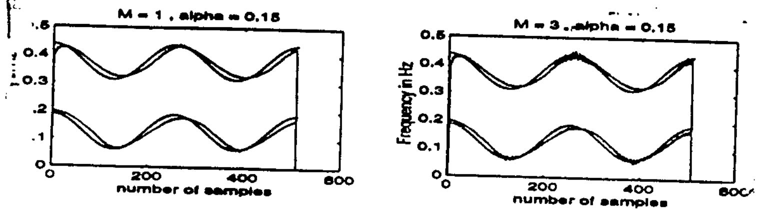


Figure 5: Adaptive tracking using (a) SLMS (b) PLMS

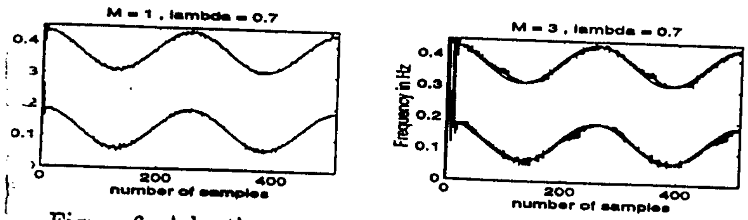


Figure 6: Adaptive tracking using (a) SRLS (b) PRLS

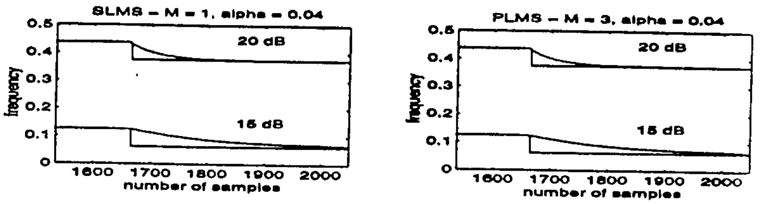


Figure 3: (a) SLMS adaptive Tracking (b) PLMS adaptive trackin

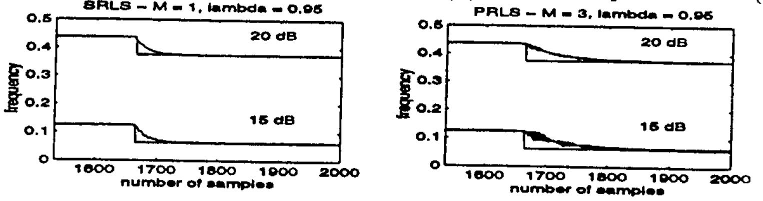


Figure 4: (a) SRLS adaptive Tracking (b) PRLS adaptive tracking

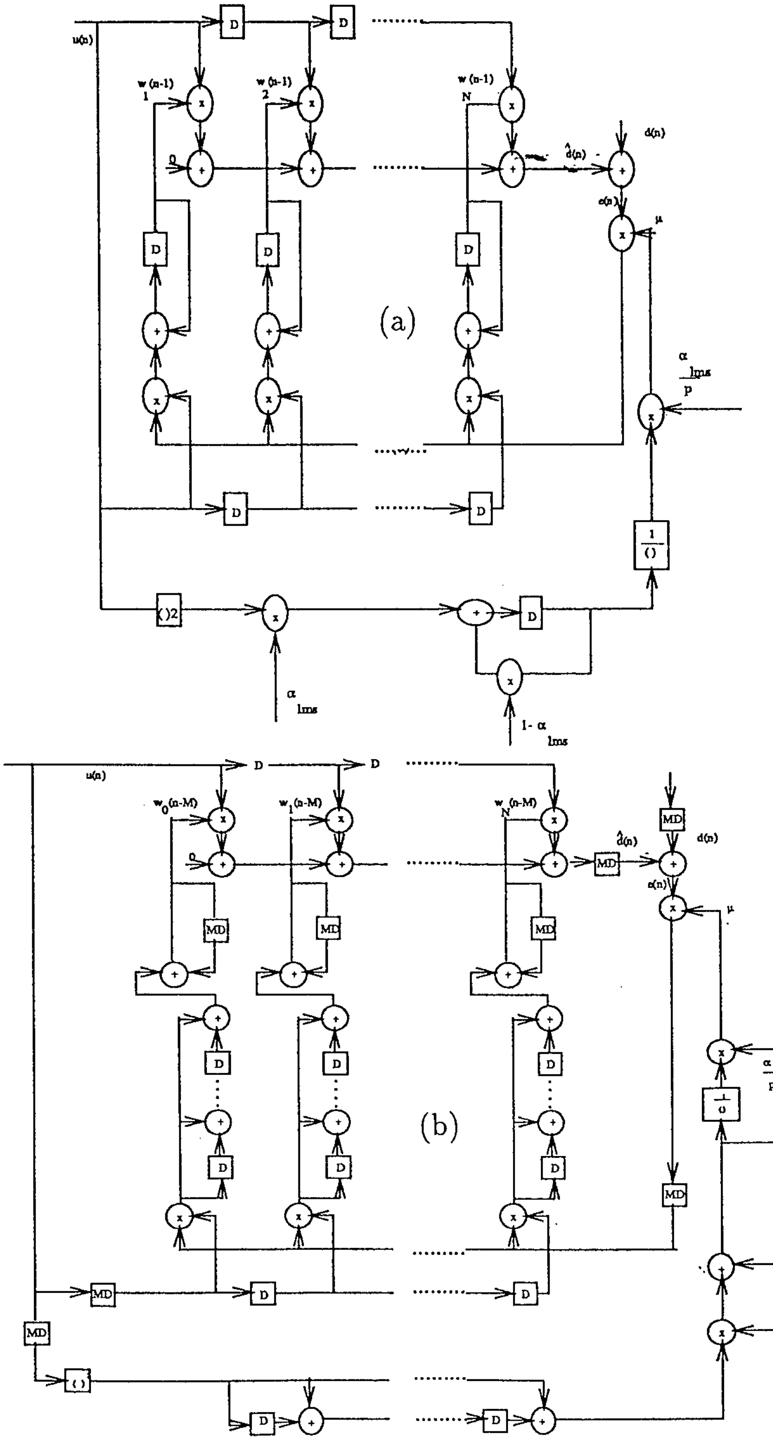


Figure 1 - (a) SLMS architecture (b) PLMS architecture

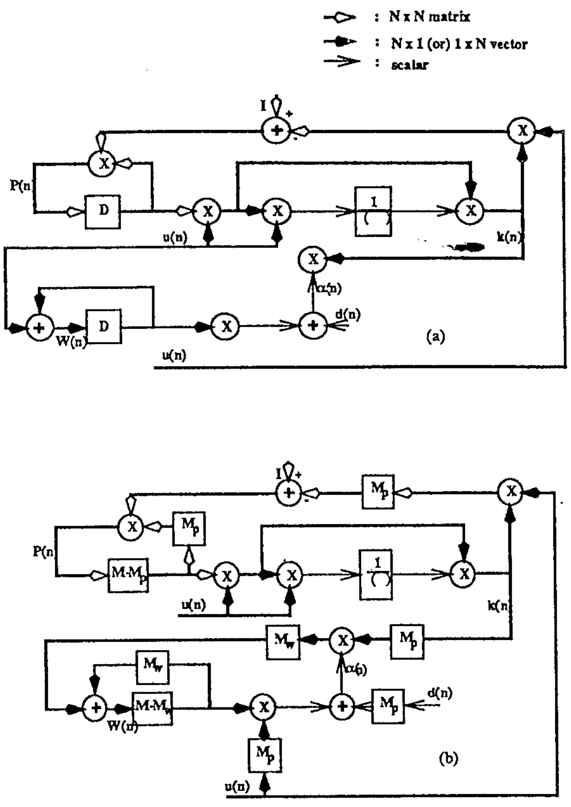


Figure 2- (a) SRLS architecture (b) PRLS architecture

References

[1] A. P. Chand, "Low Power CMOS Digital Design," *IEEE J. of Solid-State Circuits*, vol. 27, pp. 473-484, Apr., 1992.

[2] S. Haykin, "Adaptive filter Theory", Englewood Cliffs, NJ: Prentice-Hall, 2nd Ed., 1991.

[3] W. S. Hodgkiss, JR. and J. A. Presley, JR., "Adaptive Tracking of Multiple Sinusoids Whose Power Levels are Widely Separated", *IEEE Trans. Circuits Syst.*, vol. CAS-28, no. 6, pp. 550-561, June 1981.

[4] P. M. Kogge and H. S. Stone, "A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations", *IEEE Trans. Comput.*, Vol. C-22, pp786-793, Aug. 1973.

[5] Z. S. Liu, "QR Methods of $0(N)$ Complexity in Adaptive Parameter Estimation", *IEEE Trans. Signal Processing*, vol. 43, no. 3, March 1995.

[6] H.H. Loomis and B. Sinha, "High-Speed Recursive Digital Filter Realization", *Circuits, Systems and Signal Processing*, vol.3, pp. 267-294, Sept., 1984.

[7] K.K. Parhi and D.G. Messerschmitt, "Pipelining Interleaving and Parallelism in Recursive digital filters - Part I, : Pipelining using Scattered Look-Ahead and Decomposition," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, pp. 1099-1117, July 1989.

[8] K.K. Parhi , C.Y. Wang and A.P. Brown, "Synthesis of Control Circuits in Folded pipelined DSP architectures", IEEE J. of Solid-State Circuits, vol. 27, no.1, pp. 29-43, Jan. 1992.

[9] N.R. Shanbhag and K.K. Parhi, *Pipelined Adaptive Digital Filters*, Kluwer Academic Publishers, Boston, MA, 1994.

[10] M. A. Soderstrand, K. Chopper and B. Sinha, "Comparison of three new techniques for pipelining IIR digital filters," *Twenty-Third ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 439-443, Nov., 1984.

[11] J. B. Y. Tsui, *Digital Microwave Receivers : Theory and Applications*, Artech House, MA, 1989.

[12] J. F. Yang and H. J. Lin, "Adaptive High-Resolution Algorithms for Tracking Nonstationary Sources," *IEEE Transactions on Signal Processing.* vol. 42, no. 3, Jan., 1994.

CHAPTER 3

SYSTEM IDENTIFICATION AND HARWARE IMPLEMENTATION PROBLEMS

Introduction

The rational System Identification theory is closely related to the receiver design problem. In particular, Angle-of-Arrival (AOA) and frequency estimation are two of the most important integral parts of most radar receivers but these two problems can be addressed as special cases of rational system identification problems. Furthermore, digital EW receivers would require many digital filters for various purposes such as, anti-aliasing, image suppression, IF and etc.. Synthesis of digital IIR filters from any arbitrary frequency domain specifications is also one of the important problems addressed by the proposed system identification framework. Identification of unknown discrete-time linear systems is a fundamental problem in signal processing. Among many available parametric models, pole-zero or rational transfer function model is one of the most effective and practical representations. Optimal estimation of rational model parameters will be the focus of this part of the report. The system identification and signal analysis problems considered here are fundamental in nature and the results are expected to have impact and usefulness in a wide range of applications including EW receiver design.

Applications of System Identification abound in Communication systems, Automatic Control systems, Aerospace and Mechanical Systems, Econometrics and many other fields. Digital filter design from frequency and/or time-domain information has extensive applications in speech or image processing, communication, radar or sonar signal processing, bio-medical signal processing, Digital Instrumentation and Control and in various other fields. Depending on the application, the design specifications of an unknown system may be available or prescribed in the time-domain (T-D) as, (i) Impulse Response (IR) or (ii) Input-Output (IO) data, and in the frequency-domain (F-D) as (iii) Frequency Response (FR) data. The standard synthesis or identification problem is to estimate the numerator and denominator polynomial coefficients that match the prescribed specs in the least-squares (LS) sense. It is well-known that these LS problems are highly non-linear. Some existing approaches minimize 'equation errors' instead of the true fitting errors and others *modify* or *linearize* the true model-fitting criteria for iterative estimation of the numerator and denominators *simultaneously.*

The main goal in this part of the work is to exploit certain powerful theoretical results in Numerical Analysis to *theoretically decouple* the multidimensional nonlinear criteria, into two distinct problems : (1) a *purely linear* problem for estimating the numerator and (2) a non-linear problem for estimating the denominator. The nonlinear part is then reparameterized by invoking results on projection operators. In this form, the denominator criterion possesses a weighted matrix structure which is convenient for iterative optimization. But more importantly, once the optimal denominator is known, the optimal numerator is found with only a *single* step of linear LS estimation. Removal of the numerator estimation from the iterative process reduces computational complexity when compared with existing simultaneous estimators in.

The theoretical results as well as the algorithmic framework we propose here encompass a comprehensive class of system identification problems in time and frequency domains. This important underlying common theme appears to have remained unrecognized and un-utilized. In fact, one of our goal *is* to establish the analogies and equivalences between the time-domain and frequency-domain optimization approaches which seem to have evolved independently. Our hope is that a thorough study and proper understanding of these equivalences might enable us to apply and exchange useful ideas from one domain to the other. It may also lead to *combined optimization* in the frequency and time domains by matching the desired characteristics in both domains simultaneously.

The proposed unified framework is expected to provide intuitive and useful theoretical insights into various time-domain and frequency-domain identification and synthesis problems. For example, the 1-D SISO algorithms can be extended to multi-dimensional (m-D) and multi-input/multi-output (MIMO) problems in a straightforward manner.

Look-ahead pipelining has been found to be very effective for attaining high sampling rate and high computation speed in low-cost VLSI implementation of recursive digital filters. The well-known Scattered Look-ahead implementation of Recursive IIR filters achieves stability at the cost of increased multiplication and latch complexities and considerable delay in output generation. The clustered look-ahead approach can not always guarantee stability [1]. We present a new scheme (referred to as *distributed look-ahead*) which is a compromise between the two existing look-ahead approaches. The proposed scheme appears to avoid some of the potential drawbacks in various pipelined implementations of recursive filters. Our work shows that, in order to attain stability, the output samples need not be clustered or equally scattered. Indeed, in many filter design problems, stability can be maintained by using *unequally distributed* past output samples. When compared with the scattered approach, the proposed scheme uses fewer number of pole-zero cancelations and the introduced roots are not necessarily at the same radii as the original filter poles. Hence, the proposed *distributed look-ahead* scheme has reduced multiplication and latch complexities, higher area-efficiency and it produces outputs with reduced delay. The proposed DLA scheme has been used for high-speed implementation of both 1-D and 2-D Recursive Digital Filters.

The look-ahead pipelined recursive filters discussed above are obtained primarily via transformation of a *given* un-pipelined transfer function. For these approaches, it is assumed that the un-pipelined transfer function has already been designed as an intermediate step. In this project, we also present a new algorithm (OM-LA) for *direct* and *optimal* estimation of the coefficients of recursive filters in look-ahead pipelined form. OM-LA is developed by appropriate modification of a recently proposed optimal method (OM) for designing un-pipelined filters (developed previously by the PI as part of a project supported by the AFOSR). It is demonstrated that the proposed one-step approximation can achieve superior match with reduced pipelined filter order because it does not rely on pole-zero cancelations as in current LA pipelining approaches. It is also shown that the denominator polynomial can be constrained to possess any of the possible look-ahead configurations. Unlike some existing methods, OM-LA minimizes the *true* time-domain fitting error-norm between the prescribed and the estimated impulse response and produces superior results.

Section - 3.1 : Identification of 1-D Rational Systems from Input-Output Data

Summary

A theoretical and algorithmic framework is proposed for optimal identification of rational transfer function parameters of discrete-time linear systems from Input-Output (IO) data. The nonlinear criterion is theoretically *decoupled* into a purely linear problem for estimating the optimal numerator and a nonlinear problem for the optimal denominator. The proposed decoupled approach has reduced computational requirements when compared to existing algorithms which estimate the parameters simultaneously.

I. Introduction : Identification of unknown Linear Time-Invariant Discrete-Time systems is a critical problem in signal processing and control theory [1-13, 15-22]. This work addresses the problem of optimal identification of the parameters of rational transfer functions by Least-Squares (LS) fitting of observed input-output sequences. Optimization of the LS criterion for this problem requires multi-dimensional nonlinear optimization [1, 2, 15-21]. Many existing algorithms either *modify* or *linearize* the true nonlinear error criterion to estimate the unknown parameters *simultaneously.* This work will demonstrate that the optimal rational model identification problem belongs to a special class of *mixed*-nonlinear optimization framework where the linear and nonlinear variables *separate* [14]. The true nonlinear criterion will be theoretically *decoupled* into :

(i) a *purely linear* problem for obtaining the optimal numerators and

(ii) a nonlinear problem of *reduced dimensionality* for determining the optimal denominators.

The *decoupled* criteria retain the *global* optima of the original criterion. Only the criterion for the denominator is nonlinear but it possesses a weighted-matrix structure which is utilized for minimizing it iteratively. The optimal numerator is estimated in one step. Hence, unlike some existing algorithms which estimate both sets of parameters iteratively [2], the proposed computational algorithm has reduced computational requirements.

II. Problem Formulation : Rational transfer function representations of a SISO plant,

$$H(z) = \frac{a(0) + a(1)z^{-1} + \cdots + a(q)z^{-q}}{1 + b(1)z^{-1} + \cdots + b(p)z^{-p}} \triangleq \frac{A(z)}{B(z)}, \tag{1}$$

$$= h(0) + h(1)z^{-1} + \cdots + h(N-1)z^{-(N-1)} + \cdots, \tag{2}$$

where, $b(0) = 1$. Fig. 1 depicts what is commonly known as the *output-error* model of a plant, where, $y_o(n)$ and $y(n)$ denote the true and observed (possibly noisy) output signals, respectively, and $v(n)$ denotes the observation or measurement noise. Let,

$$\mathbf{x} \triangleq [x(0) \quad x(1) \quad \cdots \quad x(N-1)]^T \tag{3a}$$

and

$$\mathbf{y} \triangleq [y(0) \quad y(1) \quad \cdots \quad y(N-1)]^T \tag{3b}$$

denote the vectors containing the N input and observed samples, respectively. In vector form, the unknown model parameters are defined as,

$$\mathbf{a} \triangleq [a(0) \quad a(1) \quad \cdots \quad a(q)]^T \tag{4a}$$

and

$$\mathbf{b} \triangleq [1 \quad b(1) \quad \cdots \quad b(p)]^T. \tag{4b}$$

The problem under consideration in this part of the project can be stated as follows :

Given the observed output data **y** and the input data **x**, estimate the *optimal* model parameters **a** and **b** by minimizing the following LS model-fitting criterion :

$$\min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \left[y(i) - \frac{A(z)}{B(z)} \{x(i)\} \right]^2 . \tag{5}$$

Regarding methods related to this work, Kalman [1] had defined an equation error to solve this problem (KM), whereas Steiglitz and McBride (SMM) iteratively minimized a modified error criterion [2] to estimate both sets of parameters simultaneously.

III. PROPOSED METHOD (OM-IO) : Let $H_b(z)$ be the inverse filter corresponding to $B(z)$, *i.e.*,

$$B(z)H_b(z) \;=\; 1. \tag{6}$$

This is a convolution operation and hence, in matrix notations,

$$\mathbf{B}_b\mathbf{H}_b \;=\; \mathbf{I}_N, \tag{7}$$

where, $\mathbf{I}_N$ denotes an $N \times N$ identity matrix; $\mathbf{B}_b$ and $\mathbf{H}_b$ are convolution matrices formed as,

$$\mathbf{B}_b(i,j) \;\underset{=}{\triangle}\; b(i-j), \tag{8a}$$

and

$$\mathbf{H}_b(i,j) \;\underset{=}{\triangle}\; H_b(i-j), \qquad \text{for, } i,j \;=\; 1,\ldots,N \tag{8b}$$

Note that both these matrices are lower-triangular. In partitioned form,

$$\mathbf{B}_b \;=\; \begin{bmatrix} \mathbf{B}_u^T \\ --- \\ \mathbf{B}^T \end{bmatrix}, \qquad \mathbf{H}_b \;=\; [\,\mathbf{H}_l|\mathbf{H}_r\,]. \tag{9}$$

where, $\mathbf{B}_u \in \mathbb{R}^{N\times(q+1)}$, $\mathbf{B} \in \mathbb{R}^{N\times(N-q-1)}$, $\mathbf{H}_l \in \mathbb{R}^{N\times(q+1)}$ and $\mathbf{H}_r \in \mathbb{R}^{N\times(N-q-1)}$. Using (6) and assuming that the input is causal, *i.e.*, $x(n) \;=\; 0$, for $n < 0$, the optimization criterion in (5) can be restated as,

$$\min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \Big[y(i) - x(i) \;*\; h_b(i) \;*\; a(i) \Big]^2, \tag{10}$$

where, $*$ denotes the convolution operation. In matrix notations the problem is equivalent to :

$$\min_{\mathbf{a},\mathbf{b}} ||\mathbf{e}(\mathbf{a},\mathbf{b})||^2 \;\underset{=}{\triangle}\; \min_{\mathbf{a},\mathbf{b}} ||\mathbf{y} \;-\; \mathbf{X}\mathbf{H}_l\mathbf{a}||^2, \qquad \text{where,} \tag{11a}$$

$$\mathbf{X}(i,j) \;\underset{=}{\triangle}\; x(i-j) \qquad \text{for,} \qquad i,j \;=\; 1,\ldots,N. \tag{11b}$$

This is a mixed optimization problem where the linear and nonlinear variables appear *separately.* If $\mathbf{H}_l$ (*i.e.*, **b**) is known, then the linear LS estimate of the numerator,

$$\hat{\mathbf{a}} \;\underset{=}{\triangle}\; (\mathbf{X}\mathbf{H}_l)^{\#}\mathbf{y}, \tag{12}$$

where, $(\mathbf{X}\mathbf{H}_l)^{\#} \;\underset{=}{\triangle}\; ((\mathbf{X}\mathbf{H}_l)^T(\mathbf{X}\mathbf{H}_l))^{-1}(\mathbf{X}\mathbf{H}_l)^T$ denotes the pseudo-inverse of $(\mathbf{X}\mathbf{H}_l)$. Plugging $\hat{\mathbf{a}}$ back in (11a), the optimization criterion for **b** is given by,

$$\min_{\mathbf{b}} ||\mathbf{e}(\mathbf{b})||^2 \;\underset{=}{\triangle}\; \min_{\mathbf{b}} ||\mathbf{y} - \mathbf{P}_{\mathbf{X}\mathbf{H}_l}\mathbf{y}||^2 \;=\; \min_{\mathbf{b}} ||[\mathbf{I}_N - \mathbf{P}_{\mathbf{X}\mathbf{H}_l}]\mathbf{y}||^2 \tag{13}$$

where, $\mathbf{P}_{\mathbf{XH}_l} \triangleq \mathbf{XH}_l((\mathbf{XH}_l)^T(\mathbf{XH}_l))^{-1}(\mathbf{XH}_l)^T$, denotes the projection matrix of $(\mathbf{XH}_l)$. In (13), the parameters in $\mathbf{b}$ are indirectly related to the error criterion in a rather complicated manner through $\mathbf{P}_{\mathbf{XH}_l}$. Next, the inherent matrix-structure of the criterion in (13) is utilized to reparameterize the error criterion by relating it directly to the coefficients in $\mathbf{b}$.

Let $X_I(z)$ be the *inverse* of the input sequence $X(z)$, *i.e.*, $X(z)X_I(z) = 1$. Similar to (6) and (7), in matrix notation,

$$\mathbf{X}_I\mathbf{X} = \mathbf{I}_N, \tag{14a}$$

where, $\mathbf{X}_I \in \mathbb{R}^{N\times N}$ is also a lower triangular matrix defined as,

$$\mathbf{X}_I \triangleq x_I(i-j), \qquad \text{for}, \qquad i,j = 1,\ldots,N. \tag{14b}$$

For finite N, this inverse exists as long as the first element of the input sequence is non-zero, *i.e.*, $x(0) \neq 0$. This is not a major restriction for the causal systems under consideration in this work because the output will have non-zero leading values only when there is non-zero input. But it would be desirable that $X(z)$ be minimum-phase, otherwise $X_I(z)$ may be unbounded for some values of z which in turn may result in very high magnitudes of $x_I(n)$ for large N. Combining (7) and (14) and using the partitioned forms of (9),

$$\mathbf{B}_b\mathbf{X}_I\mathbf{X}\mathbf{H}_b = \mathbf{I}_N = \begin{bmatrix} \mathbf{B}_u^T \\ --- \\ \mathbf{B}^T \end{bmatrix} \mathbf{X}_I\mathbf{X}\,[\,\mathbf{H}_l|\mathbf{H}_r\,] \tag{15a}$$

$$\text{or,} \quad \begin{bmatrix} \mathbf{B}_u^T\mathbf{X}_I\mathbf{X}\mathbf{H}_l & | & \mathbf{B}_u^T\mathbf{X}_I\mathbf{X}\mathbf{H}_r \\ ---- & | & ---- \\ \mathbf{B}^T\mathbf{X}_I\mathbf{X}\mathbf{H}_l & | & \mathbf{B}^T\mathbf{X}_I\mathbf{X}\mathbf{H}_r \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{(q+1)} & | & \mathbf{0}_{(q+1)\times(N-q-1)} \\ -------- & | & --------- \\ \mathbf{0}_{(N-q-1)\times(q+1)} & | & \mathbf{I}_{(N-q-1)\times(N-q-1)} \end{bmatrix}. \tag{15b}$$

The bottom-left corner element shows that the $N \times (N-q-1)$ matrix $\mathbf{X}_I^T\mathbf{B}$ and the $N \times (q+1)$ matrix $\mathbf{XH}_l$ are orthogonal, *i.e.*, $(\mathbf{B}^T\mathbf{X}_I)(\mathbf{XH}_l) = \mathbf{0}_{(N-q-1)\times(q+1)}$. By construction, $rank(\mathbf{X}_I^T\mathbf{B}) + rank(\mathbf{XH}_l) = N$. Hence, according to a property of projection matrices,

$$\mathbf{P}_{\mathbf{X}_I^T\mathbf{B}} + \mathbf{P}_{\mathbf{XH}_l} = \mathbf{I}_N. \tag{16}$$

Using this result in (13), the following *reparametrized* optimization criterion is obtained,

$$\min_{\mathbf{b}} \|\mathbf{P}_{\mathbf{X}_I^T\mathbf{B}}\mathbf{y}\|^2 = \min_{\mathbf{b}} \|\mathbf{X}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{X}_I\mathbf{X}_I^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}_I\mathbf{y}\|^2 \tag{17a}$$

$$= \min_{\mathbf{b}} \mathbf{y}^T\mathbf{X}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{X}_I\mathbf{X}_I^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}_I\mathbf{y}. \tag{17b}$$

In order to obtain an expression more convenient for optimization, define,

$$\mathbf{z} \triangleq \mathbf{X}_I\mathbf{y}. \tag{18}$$

It can be easily shown that,

$$\mathbf{B}^T\mathbf{z} \triangleq \mathbf{Zb}, \tag{19a}$$

where, the matrix $\mathbf{Z}$ is constructed with the elements of $\mathbf{z}$ as,

$$\mathbf{Z} \triangleq \begin{bmatrix} z(q+1) & z(q) & \cdots & z(0) & 0 & \cdots & 0 \\ z(q+2) & z(q+1) & \cdots & z(1) & z(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(p) & z(p-1) & \cdots & \cdots & \cdots & \ldots & z(0) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(N-1) & z(N-2) & \cdots & \cdots & \cdots & \cdots & z(N-p-1) \end{bmatrix}. \tag{19b}$$

Using (19a) into (17), the optimization criterion can be re-written as,

$$\min_{\mathbf{b}} \; \mathbf{b}^T\mathbf{Z}^T(\mathbf{B}^T\mathbf{X}_I\mathbf{X}_I^T\mathbf{B})^{-1}\mathbf{Z}\mathbf{b}. \tag{20}$$

Equation (12) and the reparametrized criterion in (20) are the *final* decoupled forms. It should be emphasized here that, thus far, the theoretical derivations are mathematically exact, *i.e.,* no linearization, approximation or modification have been introduced at the outset.

According to the Theorem stated in the Appendix [14], if **b** is estimated by minimizing the criterion in (20) and if that estimate is utilized for computing **a** using (12), then the resulting estimates are the *unique* and *global* minimizers of the original criterion in (5) or (11). Furthermore, once the optimal **b** is known, the estimation of the optimal **a** in (12) is a linear problem. But more importantly, **a** needs be computed *only once.*

Algorithm : The nonlinear criterion in (20) appears to be a weighted quadratic in the unknown vector **b**. But the weight matrix $(\mathbf{B}^T\mathbf{X}_I\mathbf{X}_I^T\mathbf{B})^{-1}$ itself is dependent on the unknowns in **B**. The computational algorithm exploits this weighted quadratic structure of the criterion. At k-th iteration, the algorithm minimizes,

$$\min_{\mathbf{b}} \; \mathbf{b}^T[\mathbf{Z}^T(\mathbf{B}^{T^{(k-1)}}\mathbf{X}_I\mathbf{X}_I^T\mathbf{B}^{(k-1)})^{-1}\mathbf{Z}]\mathbf{b}, \tag{21}$$

where, $\mathbf{B}^{(k-1)}$ is formed by using the estimate of **b** obtained at the previous iteration. $\mathbf{b}^{(0)} \triangleq [1 \;\, 0 \;\, \cdots 0]^T$ can be used as the initial estimate of **b** to start the iterative process. Otherwise, the initial estimates could also be found by setting the middle matrix $(\mathbf{B}^T\mathbf{X}_I\mathbf{X}_I^T\mathbf{B})^{-1}$ to identity, *i.e.*, by optimizing,

$$\min_{\mathbf{b}} \; \mathbf{b}^T\mathbf{Z}^T\mathbf{Z}\mathbf{b}. \tag{22}$$

To ensure non-trivial solutions, $b(0)$ is set to unity. Once the iterations converge, the estimated **b** is used in (12) to *linearly* estimate the numerator coefficient vector **a**.

On the Relationships with Other Methods : The proposed theoretical and algorithmic framework appears to be the most general one in its own class of 1-D deterministic rational System Identification (SID) algorithms. In fact, a large body of work on SID can be formulated as special cases of OM-IO. For example, in case of Impulse Response (IR) fitting, *i.e.*, when $x(n) = \delta(n)$ and $y(n) \triangleq h_d(n)$, the desired IR, an optimal method (OM) has been developed recently [8, 9]. The work in this part of the project may be considered to be a further generalization OM. The Evans-Fischl Method (EFM) [5] was an early precursor of OM. But EFM dealt only with the IR fitting problem and it is applicable only for the *strictly-proper* case, *i.e.*, when, $p = q + 1$. Furthermore, the recently proposed Maximum-Likelihood Method for exponential modeling (known as, KiSS or IQML) is basically a complex version of EFM with conjugate-symmetry constraints imposed on the $B(z)$ coefficients [6, 7]. Hence, KiSS/IQML is also an important special case of OM-IO. Furthermore, when $p = q+1$, the initialization step of OM is identical to Prony's Method [10] or Covariance Method of Linear Prediction [11, 13]. For general cases, Shanks [3] and Burrus-Parks [4] also estimated the denominator using the initialization step of OM. For numerator, the linear estimator in (12) was used by Shanks whereas Burrus-Parks used, $a(k) = \sum_{i=0}^{k} b(i)h_d(k-i)$, for $k = 0, 1, \ldots, q$. Finally, the formulation presented in here appears to be quite well-suited for deconvolution [22]. Specifically, if the output and the Channel IR (or, alternately, the estimates of **a** and **b**) are available, then the criterion in (11) can be appropriately modified to obtain an LS or MLE of the unknown input vector **x**.

IV : SIMULATION RESULTS : In all figures, the true and modeled impulse responses are shown in solid and dotted lines, respectively.

Simulation 1 : In this case, white noise was passed through an ARMA(7,3) system with an arbitrary impulse response. The output was corrupted with uncorrelated white noise. The first $N = 30$ input and output samples were collected for identifying the system. True model orders were used for identifying the system. The results with 30dB and 15dB SNR values are shown in Fig. 2A and 2B, respectively.

Simulation 2 - Model Reduction : For the same data sets of Simulation-1 an ARMA(5,3) model was used for identifying the system. Note that the denominator order is less than the true order in this case. The results with 30dB and 15dB SNR values are shown in Fig. 3A and 3B, respectively.

The results of Simulation-1 indicate that the proposed algorithm is able to match the unknown model impulse response very closely by minimizing the output error norm. Simulation-2 demonstrates that the algorithm also has the capability of obtaining reduced order models with good fit.

Number of Iterations for Convergence and CPU Times : For 30dB SNR, the number of iterations for convergence for actual and reduced order cases were found to be 8 and 5, respectively. The iterations were terminated in both cases when $||\mathbf{b}_{i+1} - \mathbf{b}_i||^2 < 10^{-3}$ was achieved in each case. The corresponding CPU times on VAX-8550 were 3.0 and 2.59 seconds, respectively. Similar differences in performance were found for other SNR values also. In general, the algorithm showed rapid convergence in all simulations performed. But if the unknown system is non-minimum phase or if the SNR in the output data is too low, the algorithm may converge to a suboptimum or the estimates may oscillate. In order to guarantee convergence, the proposed iterative transformation must be a contraction mapping. This may be difficult to demonstrate in general for any arbitrary input-output data set. Theoretical analysis of the convergence properties of the iterative algorithm needs to be performed.

REFERENCES

[1] R. E. Kalman, "Design of a Self Optimizing Control System," *Trans. ASME*, vol. 80, pp. 468-478, 1958.

[2] K. Steiglitz and L.E. McBride, "A Technique for Identification of Linear Systems", *IEEE Transactions on Automatic Control*, vol. AC-10, pp. 461-464, 1965.

[3] J.L.Shanks, "Recursion Filters for Digital Processing", *Geophysics*, vol. 32, pp. 33-51, 1967.

[4] C. S. Burrus and T. W. Parks, "Time Domain Design of Recursive Digital Filters," *IEEE Transactions on Audio and Electro-Acoustics*, vol. AU-18, pp. 137-141, June, 1970.

[5] A.G. Evans and R. Fischl, "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters", *IEEE Transactions on Audio and Electro-Acoustics*, vol. AU-21, pp. 61-65, 1973.

[6] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1986.

[7] R. Kumaresan and A.K. Shaw, "Superresolution by Structured Matrix Approximation", *IEEE Transactions on Antennas and Propagation*, vol. AP-36, pp. 34-44, 1988.

[8] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," accepted for publication, *IEEE Transactions on Signal Processing*, Oct., 1991.

[9] A. K. Shaw, "An Optimal Method for Identification of Pole-Zero Transfer Functions,", *International Symposium on Circuits and Systems*, San Diego, pp. 2409-2412, May, 1992.

[10] R. Prony, "Essai Experimental et Analytique etc.," L'Polytechnique, Paris, 1 Cahier 2, pp. 24-76, 1795.

[11] L.B. Jackson, *Digital Filters and Signal Processing*, Kluwer, Boston, 1986.

[12] T. W. Parks and C. S. Burrus, *Digital Filters*, Prentice-Hall, 1987.

[13] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[14] G. H. Golub and V. Pereyra, "The Differentiation of Pseudoinverses and Nonlinear Problems Whose Variables Separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413-432, Apr., 1973.

[15] J. W. Bayless and E. O. Brigham, "Application of the Kalman Filter to Continuous Signal Restoration," *Geophysics*, vol. 35, pp. 2-23, Feb., 1970.

[16] D. Graupe, *Identification of Systems*, Huntington, New York; Littleton Education Co., 1976.

[17] P. A. Jansson, R. H. Hunt and E. K. Plyler, "Resolution Enhancement of Spectra," *Journal of the Optical Society of America*, vol. 60, pp. 596-599, May, 1970.

[18] M. Morf, G. S. Sidhu and T. Kailath, "Some New Algorithms for Recursive Estimation in Constant, Linear, Discrete-Time Systems," IEEE Transactions on Automatic Control, vol. AC-19, pp. 315-323, Aug., 1974.

[19] T. Söderström and P. Stoica, *System Identification*, Prentice Hall, NJ, 1987.

[20] L. Ljung, *System Identification: Theory for the Users*, Prentice Hall, NJ, 1987.

[21] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[22] J. M. Mendel, *Maximum-Likelihood Deconvolution*, Springer-Verlag, New York, 1990.

APPENDIX : Optimality Properties of the Separate Estimators

THEOREM - (Adapted from Theorem 2.1 in [14]) : If $\hat{\mathbf{b}}$ is a global minimizer of $||\mathbf{e}(\mathbf{b})||^2$ in (13) and $\hat{\mathbf{a}}$ is estimated using that $\hat{\mathbf{b}}$ as in (12), *i.e.*,

$$\hat{\mathbf{a}} \underline{\underline{\triangle}} (\mathbf{X}\hat{\mathbf{H}}_l)^{\#}\mathbf{y}, \qquad (A.1)$$

where, $\hat{\mathbf{H}}_l$ is formed using $\hat{\mathbf{b}}$, then $||\mathbf{e}(\hat{\mathbf{a}}, \hat{\mathbf{b}})||^2$ is a global minimizer of $||\mathbf{e}(\mathbf{a}, \mathbf{b})||^2$ and $||\mathbf{e}(\hat{\mathbf{a}}, \hat{\mathbf{b}})||^2 = ||\mathbf{e}(\hat{\mathbf{b}})||^2$. Conversely, if $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ is a global minimizer of $||\mathbf{e}(\mathbf{a}, \mathbf{b})||^2$, then $\hat{\mathbf{b}}$ is a global minimizer of $||\mathbf{e}(\mathbf{b})||^2$ and $||\mathbf{e}(\hat{\mathbf{b}})||^2 = ||\mathbf{e}(\hat{\mathbf{a}}, \hat{\mathbf{b}})||^2$. Finally, if there is an unique $\hat{\mathbf{a}}$ among all possible minimizing pairs of $||\mathbf{e}(\mathbf{a}, \mathbf{b})||^2$, then $\mathbf{a}$ must satisfy (A.1).

PROOF : See [14].

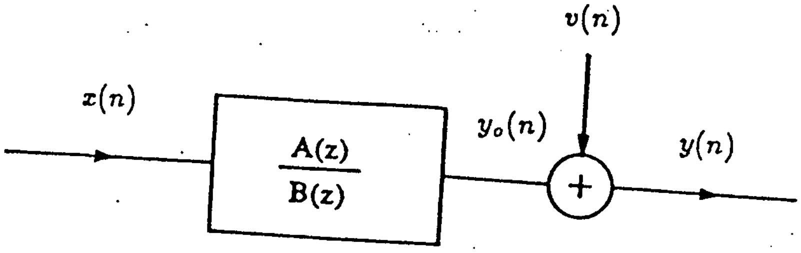


Fig. 1. The Output Error Model Structure

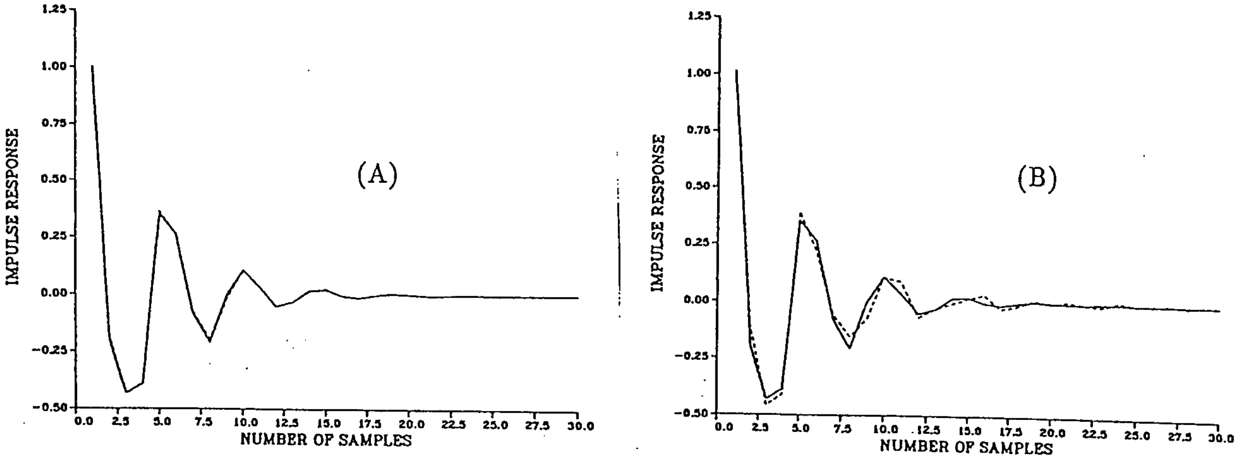


Fig. 2: Estimated Impulse Response with output SNR values of (A) 30dB and (B) 15dB. True model-order : ARMA(7,3) and assumed model-order : ARMA(7,3).

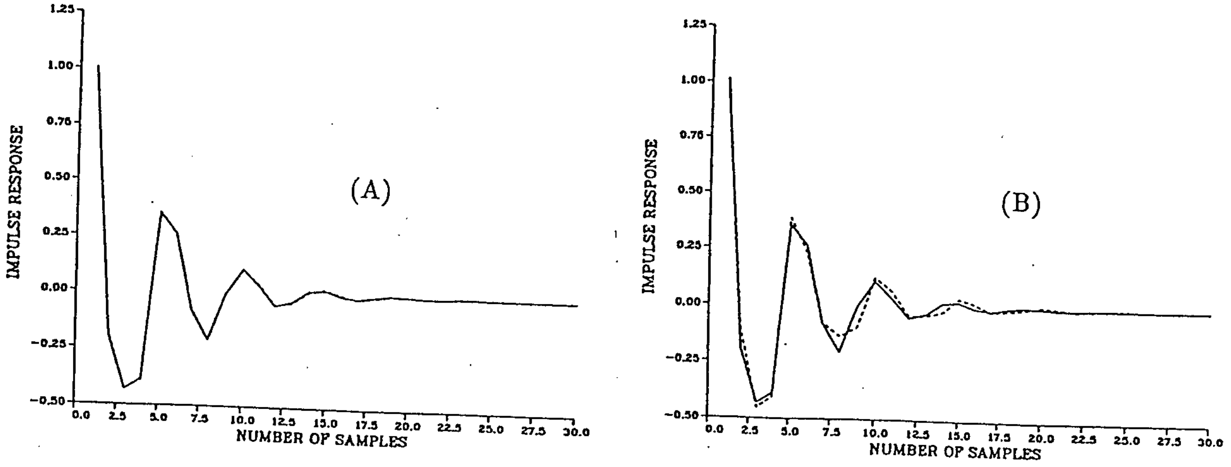


Fig. 3: Estimated Impulse Response with output SNR values of (A) 30dB and (B) 15dB. True model-order : ARMA(7,3) and assumed model-order : ARMA(5,3).

Section - 3.2 : IDENTIFICATION OF 1-D RATIONAL SYSTEMS IN THE FREQUENCY DOMAIN

SUMMARY

A new Frequency-Domain (FD) approach is presented for optimal estimation of rational transfer functions coefficients. The proposed method seeks to match any arbitrarily-shaped FD specifications in the Least-Squares (LS) sense. The desired specifications may be arbitrarily spaced in frequency. The design is performed directly in the digital domain and no analog to digital transformation is necessary. The proposed method makes use of the inherent mathematical structure in this rational modeling problem to theoretically decouple the numerator and denominator estimation problems into two smaller dimensional problems. The denominator criterion is nonlinear but possesses a weighted-quadratic structure which is convenient for iterative optimization. The optimal numerator is found linearly by solving a set of simultaneous equations. The decoupled criteria retain the global optimality properties. The performance of the algorithm is demonstrated with some simulation examples.

I : Introduction

Traditionally, digital filters are designed by performing Impulse Invariance or Bilinear transformation on available analog designs. Classical analog designs utilize polynomial approximations to match standard filter shapes such as, Low-Pass, High-Pass etc. [9, 10]. An obvious drawback of classical analog design techniques is that filters with *arbitrary* or *non-classical* specifications, as in case of a notch filter, can not be obtained. In this part of the report, a *direct* method for frequency-domain design of digital IIR filters is proposed. The method seeks to match a desired frequency response with any arbitrary shape by minimizing the optimal LS fitting error criterion. The LS criterion for this problem involves multi-dimensional nonlinear search and several linearized or modified approaches have been developed [2, 3, 21, 31]. There have been some ad-hoc attempts on designing digital filters with special shapes [9, 12]. Frequency domain version of Prony's algorithm has also been presented recently [14, 15, 19, 25]. But it appears that the underlying mathematical structure inherent in this rational modeling problem have not been fully exploited. In this work, the frequency-domain least-squares problem is formulated by identifying the orthogonal projection space which is shown to be formed entirely by the denominator parameters. The optimal denominator is estimated by minimizing the exact projection space which is independent of the numerator coefficients. The optimal numerator estimation problem turns out to be a simple linear LS problem.

It is demonstrated in this work that the optimal rational identification problem in the frequency-domain belongs to a special class of *mixed*-nonlinear optimization framework where the linear and nonlinear variables *separate* [13]. It is further shown that the true nonlinear criterion can be *decoupled* into :

(i) a *purely linear* problem for obtaining the optimal numerator coefficients and

(ii) a nonlinear problem of *reduced dimensionality* for determining the optimal denominator coefficients.

This important underlying theoretical and algorithmic aspects of designing digital filters in frequency-domain, appears to have remained mostly un-utilized. After decoupling, the denominator criterion possesses a convenient weighted-matrix structure which is then utilized to develop an iterative minimization algorithm. Once the denominator is estimated, the optimal numerator is found only once with linear LS. The *decoupled* criteria retain the *global* optima of the original criterion. The proposed approach is closely related to some time-domain results developed recently by the present author [8, 16, 17]. The design methodology described here will be based on matching desired Discrete-Time-Fourier-Transform (DTFT) values which may be arbitrarily spaced in frequency. But the algorithm can be easily modified if the desired specifications are available in the form of DFT values.

The Section is arranged as follows : In Subsection II, the rational transfer model is defined and the frequency-domain identification problem is stated. In Subsection III, some existing methods addressing this problem are briefly outlined. The details of the proposed decoupled solution is presented in Subsection IV. Some simulation examples are given in Subsection V to demonstrate the performance of the proposed approach.

II : The Rational Transfer Function Model and The Frequency-domain Design Problem

An ARMA(p, q) digital filter can be modeled as :

$$H(z) = \sum_{i=0} h(i)z^{-i} = \frac{a(0) + a(1)z^{-1} + \cdots + a(q)z^{-q}}{1 + b(1)z^{-1} + \cdots + b(p)z^{-p}} \triangleq \frac{N(z)}{D(z)}. \tag{1}$$

Let,

$$\mathbf{h} \triangleq [h(0) \quad h(1) \quad \cdots \quad h(N-1)]^T, \tag{2a}$$

be the vector with the first N significant samples of $H(z)$ and

$$\mathbf{a} \triangleq [a_0 \quad a_1 \quad \cdots \quad a_q]^T \quad \text{and} \tag{2b}$$

$$\mathbf{b} \triangleq [1 \quad b_1 \quad \cdots \quad b_p]^T \tag{2c}$$

be the numerator and denominator coefficient vectors, respectively.

Let $H_d(z)$ represent the *desired* IIR filter which needs to be modeled as $H(z)$ in (1). Using the notations of equation (1), let $H_d(\omega_k)$, $N(\omega_k)$ and $D(\omega_k)$ be defined as the frequency response values of $H_d(z)$, $N(z)$ and $D(z)$, respectively, at $z = e^{jw_k}$. The frequency-domain identification problem can be stated as follows :

> Given, $H_d(\omega_k)$, at $k = 0, 1, 2, \ldots, N-1$, the desired frequency response values (possibly arbitrarily spaced), estimate the parameters in $N(\omega_k)$ and $D(\omega_k)$ by optimizing the following LS error criterion :

$$\min_{\mathbf{a},\mathbf{b}} ||\mathbf{e}_\omega||^2 \triangleq \min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \left| H_d(\omega_i) - \frac{N(\omega_i)}{D(\omega_i)} \right|^2. \tag{3}$$

III : Some Existing Frequency-Domain Direct Design Methods

The problem stated in (3) is a nonlinear optimization problem and standard nonlinear optimization schemes can be used [7, 11]. But these generic algorithms are known to be sensitive to initial choice of estimates and they do not specifically make use of the unique mathematical structures inherent in this problem. Some linearized methods that specifically address the design problem stated in (3), have also been proposed [2, 3]. More recently, a decoupled algorithm that utilizes divided-differences and Newton-Raphson, has been reported in [14, 28]. In order to motivate the proposed algorithmic framework, brief outlines of some of the direct FD design methods are given next.

III.1 : Levy's Method (LM)

The following criterion was proposed by Levy [2] as a frequency-domain counterpart of Kalman's original work in the time-domain [1] :

$$\min_{\mathbf{a},\mathbf{b}} ||\mathbf{e}_{LM}||^2 \triangleq \min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \left| D(\omega_i)H_d(\omega_i) - N(\omega_i) \right|^2. \tag{4}$$

Note that the original error criterion in (3) is modified in Levy's case. Apart from the obvious advantage of single-step linear solution, this algorithm does not possess any other optimality properties. It may also be noted that Kalman/Levy-type approaches for the ARMA problem are closely related to Levinson's work on the all-pole problem [18], where only the first term of the error criterion was minimized. The AR parameter estimation work is further related to Prony's method [19] and Padé Approximation [20]. Similar error criterions for the ARMA problem have been later rediscovered [21] and analyzed [22].

III.2 : Sanathanan-Koerner's Prefiltering Method (SKM)

The earliest work that most closely approximates the true LS fitting-error criterion, appears to be due to Sanathanan and Koerner [3]. Their goal was to improve upon Levy's work which did not really attempt to optimize the true criterion in (3). In this case, an initial estimate of the denominator coefficients, $D^{(0)}(\omega_0)$ is first obtained by minimizing Levy's criterion in (4) and then the following *modified* fitting error criterion is optimized at the k-th iteration [3],

$$\min_{\mathbf{a},\mathbf{b}} \|\mathbf{e}_{SK}\|^2 \triangleq \min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \left| \frac{D(\omega_i)H_d(\omega_i)}{D^{(k-1)}(\omega_i)} - \frac{N(\omega_i)}{D^{(k-1)}(\omega_i)} \right|^2 . \tag{5}$$

where, $D^{(k-1)}(\omega_i)$ denotes the denominator estimate at the previous iteration which is used as a prefilter for obtaining the estimates at the following iteration step. Note that, (5) closely approximates (3) and both are identical if, $D(\omega_i) = D^{(k-1)}(\omega_i)$. But using (5), the unknown parameters in $\mathbf{a}$ and $\mathbf{b}$ can be estimated simultaneously by solving a set of linear equations. A time-domain counterpart of Sanathanan-Koerner's method was later discovered independently by Steiglitz and McBride in [4], though the later work is definitely more well-recognized in Signal Processing and System Identification literature [9, 10, 23, 24].

III.3 : Kumaresan's Decoupled Method - Generalized (KM-G)

The Frequency-Domain error criterion in (3) has been recently decoupled by Kumaresan in [14, 15, 25, 28], where *divided-difference* matrices [26] have been utilized. Similar to a time-domain decoupled algorithm due to Evans and Fischl (EFM) [6], this approach was originally proposed for strictly-proper cases, *i.e.*, when, $p = q + 1$. In the brief outline given below, appropriate modifications have been introduced in order to *generalize* KM for any arbitrary numerator and denominator orders. For q-th order numerator and p-th order denominator, the decoupled criterion for estimating the optimal denominator is :

$$\min_{\mathbf{b}} \mathbf{h}_d^{\omega H} \mathbf{C}^H (\mathbf{C}\mathbf{C}^H)^{-1} \mathbf{C} \mathbf{h}_d^{\omega} \tag{6}$$

where,

$$\mathbf{h}_d^{\omega} \triangleq [H_d(\omega_0) \quad H_d(\omega_1) \quad \cdots \quad H_d(\omega_{N-1})]^T \tag{7a}$$

denotes the vector containing the N samples of the prescribed frequency response data,

$$\mathbf{C} \triangleq \mathbf{B}_\omega^T \mathbf{U} \mathbf{D} \tag{7b}$$

$$\mathbf{B}_\omega^T \triangleq \begin{bmatrix} b(p) & \cdots & b(1) & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & b(p) & \cdots & b(1) & 1 \end{bmatrix} \in \mathbb{R}^{(N-1-q)\times(N+p-q-1)} \tag{7c}$$

$$\mathbf{U} \triangleq \begin{bmatrix} u_0^{N+p-q-2} & u_1^{N+p-q-2} & \cdots & u_{N-1}^{N+p-q-2} \\ u_0^{N+p-q-3} & u_1^{N+p-q-3} & \cdots & u_{N-1}^{N+p-q-3} \\ \vdots & \vdots & \ddots & \vdots \\ u_0 & u_1 & \cdots & u_{N-1} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{(N+p-q-1)\times N} \tag{7d}$$

$$\text{and} \quad \mathbf{D} \triangleq \begin{bmatrix} \frac{1}{\prod_{i=0,i\neq 0}^{N-1}(u_0-u_i)} & 0 & \cdots & 0 \\ 0 & \frac{1}{\prod_{i=0,i\neq 1}^{N-1}(u_1-u_i)} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \frac{1}{\prod_{i=0,i\neq N-1}^{N-1}(u_{N-1}-u_i)} \end{bmatrix} \in \mathrm{I\!R}^{N\times N} \tag{7e}$$

$$\text{with,} \qquad u_i \triangleq e^{j\omega_i}. \tag{7f}$$

Defining $\mathbf{f} \triangleq \mathbf{UDh}_d^\omega \in \mathrm{I\!R}^{(N+p-q-1)\times N}$, the error criterion can be written in the following weighted-quadratic form :

$$\min_{\mathbf{b}} \mathbf{b}^T\mathbf{F}^H(\mathbf{CC}^H)^{-1}\mathbf{Fb}, \tag{8}$$

where, $\mathbf{F} \ \in \mathrm{I\!R}^{N-q-1\times p+1}$ is formed using the elements of $\mathbf{f}$ as follows :

$$\mathbf{F} \triangleq \begin{bmatrix} f(p) & f(p-1) & \cdots & f(0) \\ f(p+1) & f(p) & \cdots & f(1) \\ \vdots & \vdots & \ddots & \vdots \\ f(N+p-q-2) & f(N+p-q-3) & \cdots & f(N-q-2) \end{bmatrix}. \tag{9}$$

The optimal denominator coefficients are obtained using an iterative algorithm. Once the optimal denominator is available, the numerator is estimated as :

$$\mathbf{a} = (\mathbf{D}_b\mathbf{U}_{q+1})^{\#}\mathbf{h}_d^\omega, \tag{10}$$

where, $^{\#}$ denotes the pseudo-inverse and

$$\mathbf{D}_b \triangleq \begin{bmatrix} \frac{1}{D(\omega_0)} & 0 & \cdots & 0 \\ 0 & \frac{1}{D(\omega_1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{D(\omega_{N-1})} \end{bmatrix} \in \mathrm{I\!R}^{N\times N} \quad \text{and} \tag{11}$$

$$\mathbf{U}_{q+1} \triangleq \begin{bmatrix} 1 & e^{j\omega_0} & e^{j2\omega_0} & \cdots & e^{jq\omega_0} \\ 1 & e^{j\omega_1} & e^{j2\omega_1} & \cdots & e^{jq\omega_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\omega_{N-1}} & e^{j2\omega_{N-1}} & \cdots & e^{jq\omega_{N-1}} \end{bmatrix} \in \mathrm{I\!R}^{N\times q+1}. \tag{12}$$

It can be easily verified that for the special case of $p = q + 1$, the general criteria given here will be exactly same as the one given in [14, 28]. It may be emphasized here that the frequency-domain LS algorithms in [2, 3] are,

(i) Approximations or modifications of the original criterion in (3), and

(ii) $(p + q)$-dimensional nonlinear optimization problems for estimating **a** and **b** simultaneously.

In contrast, the decoupled method (KM-G) estimate **a** and **b** separately. But simulation experiments indicate that the desired minimum of the criterion in (8) may not be achieved with only an Evans-Fischl type LS minimization of (8). Instead, a further step of Newton-Raphson had to be incorporated in the algorithm in order to achieve the desired optimum [14]. Unlike KM-G, the optimally decoupled method developed in this work reaches the desired optimum criterion more directly and without using Newton-Raphson.

It may be also noted that Signal Processing Toolbox of the widely popular MATLAB software package provides a direct frequency-domain design macro called `yulewalk`, which basically implements a modified Yule-Walker method developed by Friedlander and Porat [31]. This method does not attempt to minimize the true criterion in (3). Instead, it attempts to fit the deterministic correlation values to obtain the rational model parameters by essentially minimizing an equation error. The simulation section includes some comparison of the performance of the proposed method with this approach.

IV : Proposed Method (OM-DTFT)

For time-domain rational model identification problems, a new framework has been recently presented for decoupling the denominator and numerator problems into two separate but lower-dimensional optimization problems [8, 16, 17]. In this Section it is shown that the nonlinear frequency-domain criterion of (3) can also be decoupled in a similar fashion.

Let $H_b(z)$ be the inverse filter corresponding to $D(z)$, *i.e.*,

$$D(z)H_b(z) \;=\; 1. \tag{13a}$$

Clearly, this is a convolution operation in time-domain and it can be expressed using matrix notation as,

$$\mathbf{D}\mathbf{H}_b \;=\; \mathbf{I}_N, \tag{13b}$$

where, $\mathbf{I}_N$ denotes an $N \times N$ identity matrix; $\mathbf{D} \in \mathbb{R}^{N\times N}$ and $\mathbf{H}_b \in \mathbb{R}^{N\times N}$ are defined below in appropriate partitioned forms which will be useful in the algorithm :

$$\mathbf{D} \triangleq \left[\begin{array}{cccccccc} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ b(1) & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ b(q) & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline b(q+1) & \cdots & b(1) & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ b(p) & \cdots & \cdots & \cdots & b(1) & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & b(p) & \cdots & \cdots & \cdots & b(1) & 1 \end{array}\right] \triangleq \left[\begin{array}{c} \mathbf{B}_u^T \\ \hline \mathbf{B}^T \end{array}\right] \tag{14a}$$

and

$$\mathbf{H}_b \triangleq \left[\begin{array}{ccc|cc} h_b(0) & \cdots & 0 & \cdots & 0 \\ h_b(1) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_b(q+1) & \cdots & h_b(1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h_b(N-1) & \cdots & \cdots & \cdots & h_b(0) \end{array}\right] \triangleq [\mathbf{H}_l|\mathbf{H}_r]\,, \tag{14b}$$

where, $\mathbf{B}_u \in \mathbb{R}^{N\times(q+1)}$, $\mathbf{B} \in \mathbb{R}^{N\times(N-q-1)}$, $\mathbf{H}_l \in \mathbb{R}^{N\times(q+1)}$ and $\mathbf{H}_r \in \mathbb{R}^{N\times(N-q-1)}$. If the vector $\mathbf{h}$, defined in (2a), represents the finite length impulse response vector containing N significant Impulse Response values, the frequency response at any frequency ω_i will be given as,

$$H(\omega_i) \triangleq H(z)\big|_{z=e^{j\omega_i}} \;=\; \sum_{n=0}^{N-1} h(n)e^{-j\omega_i n} \tag{15}$$

Stacking the model frequency response values at all the N specified frequencies, ω_0, ω_1, $\ldots$, ω_{N-1}, the model frequency-domain vector can be expressed as :

$$\mathbf{h}_\omega \triangleq \begin{bmatrix} H(e^{j\omega_0}) \\ H(e^{j\omega_1}) \\ \vdots \\ H(e^{j\omega_{N-1}}) \end{bmatrix} \tag{16a}$$

$$= \begin{bmatrix} 1 & e^{-j\omega_0} & \ldots & e^{-j(N-1)\omega_0} \\ 1 & e^{-j\omega_1} & \ldots & e^{-j(N-1)\omega_1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\omega_{N-1}} & \ldots & e^{-j(N-1)\omega_{N-1}} \end{bmatrix} \mathbf{h} \tag{16b}$$

$$\triangleq \mathbf{Wh}. \tag{16c}$$

By definition,

$$H(z) \triangleq \frac{N(z)}{D(z)} = H_b(z)N(z), \qquad \text{using (13a)}, \tag{17}$$

where, the right-hand-side represents convolution of the numerator coefficients with the inverse sequence, $h_b(n)$ (corresponding to $H_b(z)$). Hence, it can be shown that the model impulse response vector $\mathbf{h}$ can be expressed as,

$$\mathbf{h} = \mathbf{H}_l\mathbf{a}. \tag{18}$$

Using this in (16),

$$\mathbf{h}_\omega \triangleq \mathbf{WH}_l\mathbf{a}. \tag{19}$$

With these definitions, the frequency-domain filter design problem in (3) can be restated as,

$$\min_{\mathbf{a},\mathbf{b}} \|\mathbf{e}\|^2 \triangleq \min_{\mathbf{a},\mathbf{b}} \left\|\mathbf{h}_\omega^d - \mathbf{WH}_l\mathbf{a}\right\|^2. \tag{20}$$

Equation (20) is an exact representation of the original criterion in (3), albeit in the vector-matrix form. This form of the criterion explicitly demonstrates the linear relationship between the fitting error $\mathbf{e}$ and $\mathbf{a}$ and also the nonlinear relationship between $\mathbf{e}$ and $\mathbf{b}$ through the matrix $\mathbf{H}_l$. From this equation, it is also apparent that this is a *mixed* optimization problem where the linear and nonlinear variables appear separately. In order to decouple the numerator and denominator estimation problems, consider the following. If $\mathbf{H}_l$ (*i.e.*, $\mathbf{b}$) is known, then the minimization of (20) will produce the linear LS estimate of $\mathbf{a}$ as follows,

$$\hat{\mathbf{a}} \triangleq (\mathbf{WH}_l)^{\#}\mathbf{h}_\omega^d, \tag{21}$$

where, $(\mathbf{WH}_l)^{\#} \triangleq ((\mathbf{WH}_l)^T(\mathbf{WH}_l))^{-1}(\mathbf{WH}_l)^T$. In practice though, $\mathbf{b}$ needs to be estimated also. Plugging $\hat{\mathbf{a}}$ back in (20), the optimization criterion for $\mathbf{b}$ is found as,

$$\begin{aligned} \min_{\mathbf{a},\mathbf{b}} \left\|\mathbf{h}_\omega^d - \mathbf{WH}_l\mathbf{a}\right\|^2 &\equiv \min_{\mathbf{b}} \left\|\mathbf{h}_\omega^d - \mathbf{WH}_l(\mathbf{WH}_l)^{\#}\mathbf{h}_\omega^d\right\|^2 \\ &= \min_{\mathbf{b}} \left\|\mathbf{h}_\omega^d - \mathbf{P}_{\mathbf{WH}_l}\mathbf{h}_\omega^d\right\|^2 \\ &= \min_{\mathbf{b}} \left\|(\mathbf{I}_N - \mathbf{P}_{\mathbf{WH}_l})\mathbf{h}_\omega^d\right\|^2, \end{aligned} \tag{22}$$

where, $\mathbf{P}_{\mathbf{WH}_l} \triangleq \mathbf{WH}_l((\mathbf{WH}_l)^T(\mathbf{WH}_l))^{-1}(\mathbf{WH}_l)^T$, denotes the projection matrix of $(\mathbf{WH}_l)$. Note that the numerator and denominator estimation problems are now in decoupled forms in equations (21) and (22), respectively. But in (22), the parameters in $\mathbf{b}$ are related to the error criterion in a somewhat complicated manner through $\mathbf{P}_{\mathbf{WH}_l}$. Interestingly, the operator $(\mathbf{I}_N - \mathbf{P}_{\mathbf{WH}_l})$ on $\mathbf{h}_d^\omega$ in (22) is the projection component in $\mathbf{h}_d^\omega$ that is orthogonal to the subspace spanned by the columns of $\mathbf{WH}_l$. Next it is shown that this orthogonal space can be completely defined by the denominator coefficients.

IV.1 : Reparameterization

Let $\mathbf{W}_I$ denote the inverse of the DTFT matrix $\mathbf{W}$, *i.e.*, $\mathbf{W}_I\mathbf{W} = \mathbf{I}_N$. This inverse exists as long as the frequencies ω_k's are distinct. In combination with (13b),

$$\mathbf{D}\mathbf{W}_I\mathbf{W}\mathbf{H}_b = \mathbf{I}_N. \tag{23}$$

Use of the partitioned forms of (14) into (23) leads to,

$$\begin{aligned}\left[\begin{array}{c}\mathbf{B}_u^T\\ \hline \mathbf{B}^T\end{array}\right]\mathbf{W}_I\mathbf{W}\,[\,\mathbf{H}_l|\mathbf{H}_r\,] &= \left[\begin{array}{c|c}\mathbf{B}_u^T\mathbf{W}_I\mathbf{W}\mathbf{H}_l & \mathbf{B}_u^T\mathbf{W}_I\mathbf{W}\mathbf{H}_r\\ \hline \mathbf{B}^T\mathbf{W}_I\mathbf{W}\mathbf{H}_l & \mathbf{B}^T\mathbf{W}_I\mathbf{W}\mathbf{H}_r\end{array}\right]\\ &= \left[\begin{array}{c|c}\mathbf{I}_{(q+1)} & \mathbf{0}_{(q+1)\times(N-q-1)}\\ \hline \mathbf{0}_{(N-q-1)\times(q+1)} & \mathbf{I}_{(N-q-1)\times(N-q-1)}\end{array}\right].\end{aligned} \tag{24}$$

The bottom-left corner element shows that the $N \times (N-q-1)$ matrix $\mathbf{W}_I^T\mathbf{B}$ and the $N \times (q+1)$ matrix $\mathbf{WH}_l$ are orthogonal, *i.e.*, $(\mathbf{B}^T\mathbf{W}_I)(\mathbf{WH}_l) = \mathbf{0}_{(N-q-1)\times(q+1)}$. By construction,

$$rank(\mathbf{W}_I^T\mathbf{B}) + rank(\mathbf{WH}_l) = N. \tag{25a}$$

Hence, using a property of projection matrices,

$$\mathbf{P}_{\mathbf{W}_I^T\mathbf{B}} + \mathbf{P}_{\mathbf{WH}_l} = \mathbf{I}_N. \tag{25b}$$

Using this result in (22),

$$\min_{\mathbf{b}} \ ||\mathbf{e}_b||^2 \overset{\triangle}{=} \min_{\mathbf{b}} \ ||\mathbf{P}_{\mathbf{W}_I^T\mathbf{B}}\mathbf{h}_\omega^d||^2 \tag{26a}$$

$$= \min_{\mathbf{b}} \ ||\mathbf{W}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{W}_I\mathbf{h}_\omega^d||^2 \tag{26b}$$

$$= \min_{\mathbf{b}} \ \mathbf{h}_\omega^{d^T}\mathbf{W}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{W}_I\mathbf{h}_\omega^d. \tag{26c}$$

Note that this *reparameterized* criterion is directly related to $\mathbf{b}$, as desired. In order to further simplify this expression, define a vector $\mathbf{z}$ of length N as,

$$\mathbf{z} \overset{\triangle}{=} \mathbf{W}_I\mathbf{h}_\omega^d \tag{27}$$

such that the criterion in (26) becomes,

$$\min_{\mathbf{b}} \ \mathbf{z}^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{z}. \tag{28}$$

It can be easily shown that,

$$\mathbf{B}^T\mathbf{z} \overset{\triangle}{=} \mathbf{Z}\mathbf{b}, \tag{29a}$$

where, the matrix $\mathbf{Z}$ is constructed with the elements of $\mathbf{z}$ as,

$$\mathbf{Z} \underline{\underline{\triangle}} \begin{bmatrix} z(q+1) & z(q) & \cdots & z(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(p) & z(p-1) & \cdots & \cdots & \cdots & z(1) & z(0) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z(N-1) & z(N-2) & \cdots & \cdots & \cdots & \cdots & z(N-p-1) \end{bmatrix}. \tag{29b}$$

Using (29) in (28), the optimization criterion can be rewritten as,

$$\min_{\mathbf{b}} \ \mathbf{b}^T \mathbf{Z}^T (\mathbf{B}^T \mathbf{W}_I \mathbf{W}_I^T \mathbf{B})^{-1} \mathbf{Z} \mathbf{b}. \tag{30}$$

Note that this alternate form has a weighted-quadratic structure which is convenient for minimization. Equations (21) and (30) represent the final *decoupled* estimators to be utilized in the algorithm described below. It should be emphasized here that, thus far, the theoretical derivations are mathematically exact, *i.e.*, no *linearization, approximation* or *modification* have been introduced at the outset.

Regarding optimality properties of the decoupled estimators, theoretical results in [13] can be used to prove that if $\mathbf{b}$ is estimated by minimizing the criterion in (30) and if that estimate is utilized for computing $\hat{\mathbf{a}}$ using (21), then the resulting estimates are the *unique* and *global* minimizers of the criterion in (20). The advantage of estimating the linear and nonlinear parameters independently is reduction in computational load because the iterative part is only with respect to the p coefficients in b. Based on the optimal $\mathbf{b}$, estimation of the optimal $\mathbf{a}$ is a simple linear least squares problem. But more importantly, $\mathbf{a}$ needs to be computed *only* once.

IV.2 : Algorithm

The nonlinear optimization criterion in (30) possesses a very useful matrix structure. Specifically, the expression appears to be a weighted quadratic criterion in the unknown vector $\mathbf{b}$. The matrices $\mathbf{Z}$ and $\mathbf{W}_I$ are known. But the weight matrix $(\mathbf{B}^T \mathbf{W}_I \mathbf{W}_I^T \mathbf{B})^{-1}$ itself is dependent on the unknowns in $\mathbf{B}$. The computational algorithm will utilize this weighted quadratic structure of the criterion to formulate the iterations. Specifically, the algorithm minimizes the following quadratic error criterion at k-th iteration step :

$$\min_{\mathbf{b}} \ \mathbf{b}^T [\mathbf{Z}^T (\mathbf{B}^{T^{(k-1)}} \mathbf{W}_I \mathbf{W}_I^T \mathbf{B}^{(k-1)})^{-1} \mathbf{Z}] \mathbf{b} \ \underline{\underline{\triangle}} \ \min_{\mathbf{b}} \ \mathbf{b}^T \mathbf{R}_1 \mathbf{b} \tag{31a}$$

where, $\mathbf{B}^{(k-1)}$ is formed by using the estimate of $\mathbf{b}$ obtained at the previous iteration and $\mathbf{R}_1 \ \underline{\underline{\triangle}} \ [\mathbf{Z}^T (\mathbf{B}^{T^{(k-1)}} \mathbf{W}_I \mathbf{W}_I^T \mathbf{B}^{(k-1)})^{-1} \mathbf{Z}]$ is the weight-matrix. An initial estimate of $\mathbf{b}$ is necessary to start the iterative process. $\mathbf{b}^{(0)} \ \underline{\underline{\triangle}} \ [1 \ 0 \ \cdots 0]^T$ can be used or the initial estimates could also be found setting the middle matrix $(\mathbf{B}^T \mathbf{W}_I \mathbf{W}_I^T \mathbf{B})^{-1}$ to identity, *i.e.*, by optimizing,

$$\min_{\mathbf{b}} \ \mathbf{b}^T \mathbf{Z}^T \mathbf{Z} \mathbf{b} \ \underline{\underline{\triangle}} \ \min_{\mathbf{b}} \ \mathbf{b}^T \mathbf{R}_2 \mathbf{b} \tag{31b}$$

where, the weight-matrix $\mathbf{R}_2 \ \underline{\underline{\triangle}} \ \mathbf{Z}^T \mathbf{Z}$. In order to ensure non-trivial solutions, the first term of the denominator, $b(0)$ is set to unity. The computational algorithm is similar in nature to the time-domain counterparts developed recently [8, 16, 17]. As outlined in the Appendix, the algorithm has two phases. In Phase-1, the criterion in (31a) is minimized by neglecting the variation *w.r.t.* the weight matrix. Simulation experience shows that this Phase alone brings the error quite close to the minimum. But if necessary, the variation of the weight matrix may also be included by invoking Phase-2, where the gradient of the entire criterion is set to zero. Once the iterations converge, the estimated $\mathbf{b}$ is used in (21) to *linearly* estimate the numerator coefficient vector $\mathbf{a}$.

V : Simulation Results

Two examples are included to demonstrate the effectiveness of the proposed algorithm. The first example considers a Lowpass filter design problem whereas the second one designs a Notch filter. In all plots the frequency response values are displayed up to half the sampling frequency. For the proposed method, only the Phase-1 results are given.

Simulation-1 : Lowpass Filter Design

Magnitude response values at 56 frequency points around the unit circle were taken for the matching purpose. In Fig. 1 the estimated response with $p = 6$ and $q = 5$ for the proposed method are shown by the dashed curve and the solid line represents the desired response. The algorithm converged in 6 iterations. For the sake of comparing with a widely used direct method, the Modified Yule-Walker method [30, 31] available in the MATLAB software package was used to design a 6th order filter. The magnitude response fit for this case is shown as the dot-dash line in Fig. 1.

Simulation-2 : Notch Filter

A Notch Filter design problem was considered in this case. The magnitude response values at 101 frequency points around the unit circle were taken. The estimated response with 10th order denominator and 9th order numerator as produced by the proposed method as well as the desired response are shown in dB scale in Fig. 2 in dashed and solid lines, respectively. The algorithm converged in 11 iterations. The dash-dot line again shows the fit when the Modified Yule-Walker method [30, 31] was used to design the 10th order filter.

Discussion

The first example has been adopted from [14, 28]. The results presented above for the proposed method did not have to make use of any generic nonlinear optimization technique, such as Newton-Raphson to reach the final optimum. Also, during the minimization process, all the coefficients were enforced to be real and hence the filter is readily realizable. It may also be stated here that the final designs were stabilized using the macro called Polystab available in MATLAB [29, 30], where the unstable roots are flipped inside the unit circle. The simulations clearly demonstrate that the proposed method can closely match arbitrarily shaped frequency response data and it also appears to perform better than a widely used method for direct design.

References

[1] R. E. Kalman, "Design of a Self Optimizing Control System," *Trans. ASME*, Vol. 80, pp. 468-478, 1958.

[2] E. C. Levy, "Complex Curve Fitting," *IRE Transactions on Automatic Control*, vol. AC-4, pp.-37-44, May, 1959.

[3] C. K. Sanathanan and J. Koerner, "Transfer Function Synthesis as a Ratio of Two Complex Polynomials," *IEEE Transactions on Automatic Control*, vol. AC-8, pp. 56-58, January, 1963.

[4] K. Steiglitz and L.E. McBride, "A Technique for Identification of Linear Systems", *IEEE Transactions on Automatic Control*, Vol. AC-10, pp. 461-464, 1965.

[5] C. S. Burrus and T. W. Parks, "Time Domain Design of Recursive Digital Filters," *IEEE Trans. on Aud. Elect.*, vol. AU-18, pp. 137-141, June, 1970.

[6] A.G. Evans and R. Fischl, "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters", *IEEE Transactions on Audio and Electro-Acoustics*, Vol. AU-21, pp. 61-65, 1973.

[7] J. A. Cadzow, "Recursive Digital Filter Synthesis via Gradient Based Algorithms", *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. ASSP-24, pp. 349-355, 1976.

[8] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," *IEEE Trans. on Signal Processing*, Jan., 1994.

[9] L.B. Jackson, *Digital Filters and Signal Processing*, Kluwer, Boston, 1986.

[10] T. W. Parks and C. S. Burrus, *Digital Filters*, Prentice-Hall, 1987.

[11] R. Fletcher and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, Vol. 6, pp. 163-168, 1963.

[12] A. Nehorai, "A Minimal Parameter Adaptive Notch Filter With Constrained Poles and Zeros," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-33, no. 4, pp. 983-996, Aug., 1985.

[13] G. H. Golub and V. Pereyra, "The Differentiation of Pseudoinverses and Nonlinear Problems Whose Variables Separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413-432, Apr., 1973.

[14] R. Kumaresan and C. S. Burrus, "Fitting a Pole-Zero Filter Model to Arbitrary Frequency Response Samples," *ASILOMAR-91*, pp. 1649-1652, 1991.

[15] R. Kumaresan, "On the Frequency-Domain Analog of Prony's Method," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-38, no. 1, pp. 168-170, Jan., 1990.

[16] A. K. Shaw, "An Optimal Method for Identification of Pole-Zero Transfer Functions," *International Symposium on Circuits and Systems*, San Diego, pp. 2409-2412, May, 1992.

[17] A. K. Shaw, "A Decoupled Approach for Optimal Estimation of Transfer Function Parameters from Input-Output Data," under review, *IEEE Transactions on Acoustics, Speech and Signal Processing.*

[18] N. Levinson, "The Wiener RMS Error Criterion in Filter Design and Prediction," *J. Math. Phys.*, vol. 25, pp. 261-278, Jan. 1947.

[19] R. Prony, "Essai Experimental et Analytique etc.," L'Polytechnique, Paris, 1 Cahier 2, pp. 24-76, 1795.

[20] L. Weiss and R. N. McDonough, "Prony's Method, *Z*-Transforms, and Padé Approximations," *SIAM Review*, vol. 5, no. 2, April, 1963.

[21] J. N. Brittingham, E. K. Miller and J. L. Willows, "Pole Extraction from Real Frequency Information," *IEEE Proceedings*, vol. 68, pp. 263-273, Feb, 1980.

[22] C. T. Mullis and R. A. Roberts, "The Use of Second-Order Information in the Approximation of Discrete-Time Linear Systems," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-24, no. 3, pp. 226-238, June, 1976.

[23] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[24] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, NJ, 1992.

[25] R. Kumaresan, "On Identifying a Rational Transfer Function from the Frequency Response Samples", *IEEE Trans. on Aerospace and Electronic Systems*, pp. 925, Nov. 1990.

[26] F. B. Hildebrand, *Introduction to Numerical Analysis*, 2nd Edition, Dover Publications, New York, 1987.

[27] R. Kumaresan, "Parameter Estimation of Signals Corrupted by Noise Using a Matrix of Divided Differences," Invited Paper presented at the NATO Advanced Studies Institute on, *Underwater Acoustic Data Processing,*

Canada, July, 1988.

[28] J. Romero, *Fitting a Pole-Zero Model to Arbitrary Frequency Response Samples*, MS Thesis, University of Rhode Island, Kingston, Rhode Island, 1989.

[29] MATLAB 4.0, Reference Guide, *The MathWorks, Inc.*, Natick, Massachusetts, 1993.

[30] Signal Processing Toolbox for use with MATLAB, User's Guide, *The MathWorks, Inc.*, Natick, Massachusetts, 1992.

[31] B. Friedlander and B. Porat, "The Modified Yule-Walker Method of ARMA Spectral Estimation," *IEEE Trans. on Aerospace and Electronic Systems*, no. 2, pp. 158-173, Mar., 1984.

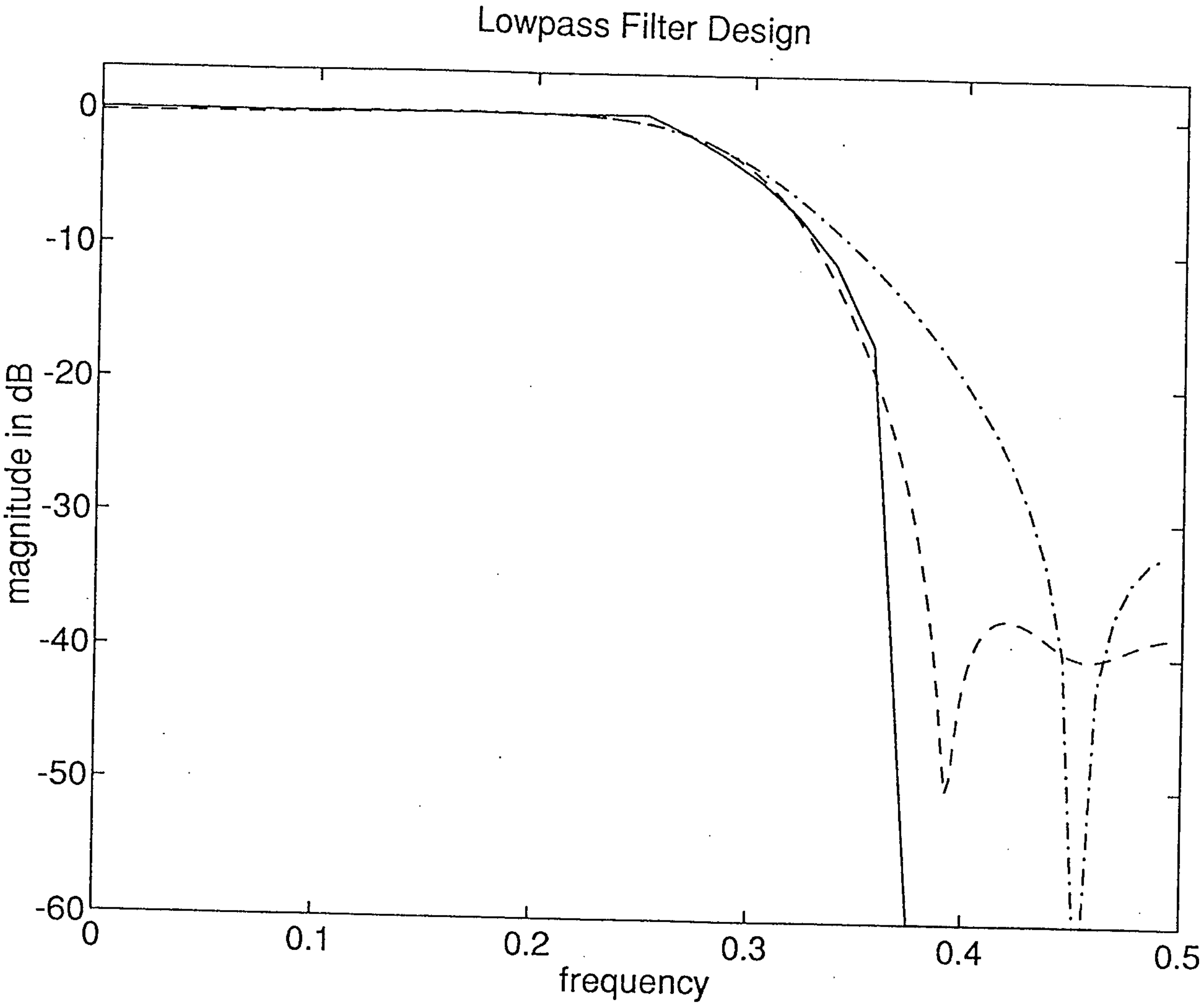


Fig. 1 : The desired Lowpass response is shown as the solid line. The estimated responses using the proposed method and the Yule-Walker method are shown in dashed and dot-dash lines, respectively. The filter order is six.

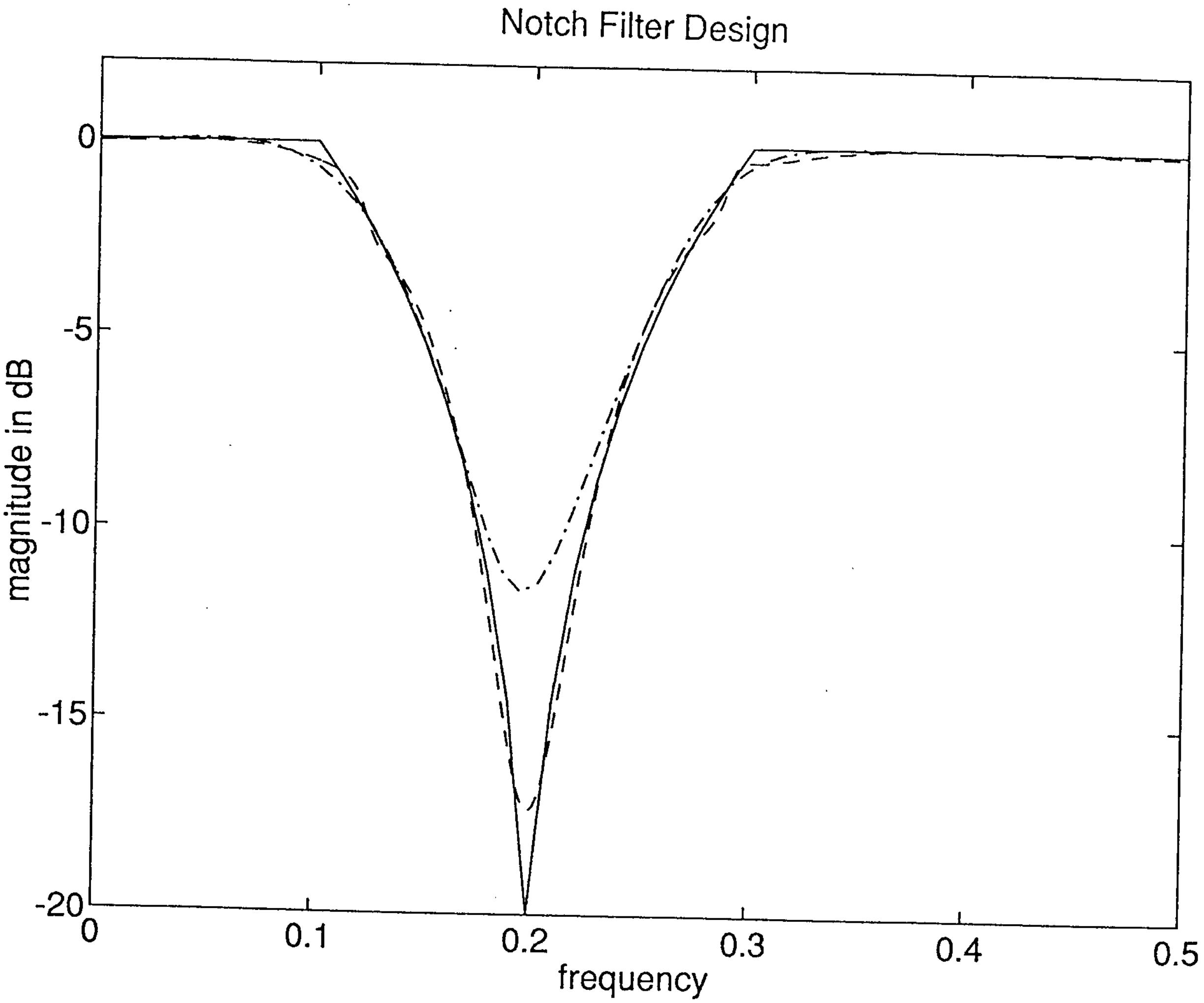


Fig. 2 : The desired Notch Filter is shown as the solid line. The estimated responses using the proposed method and the Yule-Walker method are shown in dashed and dot-dash lines, respectively. The filter order is ten.

APPENDIX : Computational Algorithm

The criterion in (31a) is non-linear in $\mathbf{b}$ and hence it can not be minimized directly. But instead of using any generic non-linear optimization techniques, the inherent mathematical structure of the criterion will be utilized to develop an iterative computational algorithm. The algorithm consists of two phases. In Phase-1, the variations in the middle matrix $(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})$ in (31a) is not taken into account in the derivative calculations, whereas in Phase-2 the gradient of the error norm in (31a) is set to zero.

Phase-1

The final form of the error vector in (31a) is rewritten as,

$$\begin{aligned}
\mathbf{e}_b &= \mathbf{W}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}\mathbf{Z}\mathbf{b} \qquad (A.1)\\
&\triangleq \mathbf{V}\mathbf{Z}\mathbf{b} \qquad (A.2)\\
&\triangleq \mathbf{V}\left[\mathbf{g} \,\vdots\, \mathbf{G}\right]\mathbf{b} \qquad (A.3)\\
&= \mathbf{V}\mathbf{g} + \mathbf{V}\mathbf{G}\hat{\mathbf{b}}, \qquad (A.4)
\end{aligned}$$

where,

$$\mathbf{V} \triangleq \mathbf{W}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}, \tag{A.5}$$

and

$$\hat{\mathbf{b}} \triangleq [b(1)\ b(2)\ \ldots\ b(p)]^T. \tag{A.6}$$

If the matrix $\mathbf{V}$ is treated as independent of $\hat{\mathbf{b}}$, an expression for $\hat{\mathbf{b}}$ can be easily obtained by minimizing $||\mathbf{e}_b||^2$ *w.r.t.* $\hat{\mathbf{b}}$ as follows :

$$\begin{aligned}
\hat{\mathbf{b}} &= -(\mathbf{V}\mathbf{G})^{\#}\mathbf{V}\mathbf{g}\\
&= -\left(\mathbf{G}^T\mathbf{V}^T\mathbf{V}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{V}^T\mathbf{V}\mathbf{g}. \qquad (A.7)
\end{aligned}$$

But since $\mathbf{V}$ does depend on the elements in $\hat{\mathbf{b}}$, (A.7) can only be computed iteratively. At the $(i+1)$-th step of iteration, $\mathbf{V}^{(i)}$ is formed using the estimate of $\mathbf{b}$ found in the i-th iteration step. This leads to the following iterative algorithm for computing $\mathbf{b}^{i+1}$:

$$\mathbf{b}^{(i+1)} = \begin{bmatrix} 1 \\ \ldots\ldots\ldots\ldots\ldots \\ -[\mathbf{F}^{(i)}\mathbf{G}]^{-1}[\mathbf{F}^{(i)}]\mathbf{g} \end{bmatrix} \tag{A.8}$$

where,

$$\mathbf{F}^{(i)} \triangleq \mathbf{G}^T\mathbf{V}^{T^{(i)}}\mathbf{V}^{(i)} \tag{A.9}$$

The iterations are continued until $||\mathbf{b}_{i+1} - \mathbf{b}_i||^2 < \delta$, where δ is an arbitrarily small number. It must be noted here that the iterations in (A.8) may not always converge to the absolute minimum of the error criterion in (31a) and hence the estimated $\mathbf{b}$ may not be the optimum one. This is because in (A.8) the variability of $\mathbf{V}$ *w.r.t.* $\mathbf{b}$ had been ignored while minimizing $||\mathbf{e}||^2$. To achieve the optimum, the gradient of the complete expression of $||\mathbf{e}||^2$ must be set to zero. If desired, this can be done in Phase-2 of the algorithm which is outlined next. It may be noted here that the simulation studies indicate that the Phase-1 of iterations using (A.8) perform an excellent job of bringing the estimate very close to the optimum. Once the estimates of $\mathbf{b}$ converge, $\mathbf{a}$ is computed using (21).

Phase-2

In this phase, the derivative of the matrix $\mathbf{V}$ *w.r.t.* $\hat{\mathbf{b}}$ is taken into consideration while minimizing the fitting error norm. By setting the derivative of the squared norm in (31a) to zero, we obtain the updated $\hat{\mathbf{b}}^{(i+1)}$ at the $(i+1)$-th iteration as,

$$\hat{\mathbf{b}}^{(i+1)} = -[\mathbf{S}^{(i)}\mathbf{G}]^{-1}[\mathbf{S}^{(i)}]\mathbf{g} \tag{$A.10$}$$

where (suppressing the superscript $^{(i)}$),

$$\mathbf{S} \triangleq \mathbf{L}^T\mathbf{V} + \mathbf{G}^T\mathbf{V}^T\mathbf{V}, \tag{$A.11a$}$$

$$\mathbf{L} \triangleq \left[\frac{\partial \mathbf{V}}{\partial b(1)}\mathbf{Zb} \;\middle|\; \cdots \middle| \frac{\partial \mathbf{V}}{\partial b(p)}\mathbf{Zb}\right], \tag{$A.11b$}$$

$$\begin{aligned}\frac{\partial \mathbf{V}}{\partial b(k)} \triangleq \frac{\partial}{\partial b(k)}[\mathbf{W}_I^T\mathbf{B}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}] &= \mathbf{W}_I^T\frac{\partial \mathbf{B}}{\partial b(k)}(\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1} - \mathbf{W}_I^T\mathbf{B}\left[\left[\frac{\partial \mathbf{B}^T}{\partial b(k)}\right]\mathbf{W}_I\mathbf{W}_I^T\mathbf{B}\right.\\ &\left. + \mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\left[\frac{\partial \mathbf{B}}{\partial b(k)}\right]\right](\mathbf{B}^T\mathbf{W}_I\mathbf{W}_I^T\mathbf{B})^{-1}\end{aligned} \tag{$A.11c$}$$

and $\frac{\partial \mathbf{B}}{\partial b(k)}$ has the same form as the $\mathbf{B}$ matrix defined in (10) but it is filled with all zeros except at the locations where $b(k)$ appears. For example,

$$\frac{\partial \mathbf{B}}{\partial b(p)} = \begin{bmatrix} 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \cdots & 0 & 0 & \ddots & 1 \\ 0 & \cdots & 0 & 0 & \ddots & 0 \\ \vdots & \cdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{N \times N-q-1} \tag{$A.12$}$$

Once $\hat{\mathbf{b}}^{(i+1)}$ is found using (A.10), $\mathbf{b}^{(i+1)}$ can be formed as,

$$\mathbf{b}^{(i+1)} = \begin{bmatrix} 1 \\ \cdots\cdots \\ \hat{\mathbf{b}}^{(i+1)} \end{bmatrix} \tag{$A.13a$}$$

$$= \begin{bmatrix} 1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ -[\mathbf{S}^{(i)}\mathbf{G}]^{-1}[\mathbf{S}^{(i)}]\mathbf{g} \end{bmatrix}. \tag{$A.13b$}$$

This minimization phase continues until $\mathbf{b}^{i+1} \simeq \mathbf{b}^i$ is reached and this optimum $\mathbf{b}$ vector corresponds to a minimum of the error surface of $||\mathbf{e}_b||_2^2$.

Section - 3.3 : Optimal Estimation of the Parameters of All-Pole Transfer Functions

Summary

An algorithm is proposed for optimal estimation of the parameters of Auto-Regressive (AR) or all-pole transfer function models from prescribed impulse response data. The transfer function coefficients are estimated by minimizing the ℓ_2-norm of the exact model fitting error. Existing methods either minimize equation errors or modify the true non-linear fitting error criterion. In the proposed method, the multidimensional nonlinear error criterion has been decoupled into a purely linear and a nonlinear subproblem. Global optimality properties of the decoupled estimators have been established. For data corrupted with Gaussianly distributed noise, the proposed method produces Maximum-Likelihood Estimates (MLE) of the AR-parameters. The inherent mathematical structure in the non-linear subproblem is exploited in formulating an efficient iterative computational algorithm for its minimization. The proposed algorithm provides an useful computational tool based on appropriate theoretical foundation for accurate modeling of all-pole systems from prescribed impulse response data. The effectiveness of the algorithm has been demonstrated with several simulation examples.

1. Introduction

Parameter estimation of unknown discrete-time linear systems is a fundamental problem in digital signal processing. Parametric models overcome the infinite dimensionality problem of non-parametric models with parsimonious representation of systems in terms of only a finite number of parameters. Over the last few decades these problems have been addressed in a large body of work in many different fields [1-17, 22, 24-38, 40-47]. Among many parametric models used in signal processing, Auto-Regressive (AR) or all-pole model is one of the most effective and practical representations.

The AR-parameter identification problem arises both in stochastic and deterministic time-series analysis. There are probably two primary reasons for the wide popularity of AR modeling in statistical time series analysis. Firstly, according to Kolmogorov Theorem, any minimum phase transfer function $H(z)$ can be represented by a possibly infinite order, stable minimum phase AR-model [9, 10]. This important theorem implies that even if an AR model is picked erroneously, the unknown Power Spectral Density can still be matched closely as long as a 'large enough' AR model order is chosen. But the second and the main reason for the popularity of AR-modeling is that it is possible to obtain reasonably good *suboptimal* estimates of the unknown AR-parameters by solving a simultaneous set of linear equations.

Modeling human vocal-tracts as all-pole systems and the corresponding Speech signal as AR-process is one of the most important applications of AR-modeling [4, 8]. Furthermore, two important modeling philosophies, viz., Linear Prediction (LP) and Maximum Entropy methods essentially produce the AR-parameters as their estimates, regardless of the true underlying signal model.

This Section deals with the problem of estimating the parameters of an all-pole transfer function to match a prescribed or desired impulse response specification. The least-squares Impulse Response (IR) model fitting error has been chosen as the objective optimality criterion. Many well-known techniques developed for statistical time-series analysis have been used successfully in the deterministic case also [7, 10]. AR-model fitting may be considered a special case of estimating the unknown parameters of general ARMA (or pole-zero) models. ARMA parameter estimation is known to be a multidimensional nonlinear optimization problem and there have been extensive work on this subject [1-7, 10, 12, 13, 15, 16, 24-31, 34-37, 41-43, 47]. In one of the earlier works, Kalman [1] had proposed a linearized and approximate 'equation error' minimization technique which produces suboptimal estimates. Several two-step procedures where the denominator and numerator polynomials are estimated separately, have also been proposed [2, 30]. In these methods, the denominator is first estimated by minimizing an 'equation error' and then the numerator is found by minimizing a linearized 'fitting error' [2] or by setting the leading error samples to zeros [30]. A thorough coverage on filter design by modeling may be found in [7].

Equation error minimization is a commonly used optimization procedure for estimating AR-parameters. In fact, the well-known linear prediction (LP) coefficients [7-10] are estimated by minimizing equation errors. Linear predictors have extensive usefulness in speech analysis, synthesis and coding. Many practical and efficient algorithms are available for LP parameter estimation. Among these, the 'Autocorrelation Method' (AM) and the 'Covariance Method' (CM) are most popular. CM and AM do not produce optimal estimates in the sense that the *model fitting error* norm is not minimized in either case. In contrast, Steiglitz and McBride (SMM) had proposed a *modified* fitting error minimization criterion for estimating the coefficients of general ARMA models [3, 4, 7]. SMM has also been adapted for AR parameter estimation [7]. In absence of any exact model fitting error criterion, SMM has established itself as the standard method for AR and ARMA parameter estimation problems [3, 4, 7, 10, 12, 22, 25, 34, 40, 47]. In [5], a decoupled exact fitting error minimization approach has also been proposed by Evans and Fischl (EFM). But their algorithm is applicable *only* in the case of strictly proper ARMA models where the number of poles must be exactly one more than the number of zeros. Consequently, the optimal EFM can be applicable for identifying first-order AR-models only. The proposed optimal algorithm has no such restrictions.

The proposed algorithm originates from a recently developed optimal method (OM) for general ARMA modeling [6]. Unlike EFM, the decoupled fitting error minimization approach in [6] is applicable for ARMA models with arbitrary numbers of poles and zeros. Furthermore, in contrast to the methods in [1-4, 24, 30], no linearization or modification of error criterion is introduced in the theoretical derivation of the least-squares model fitting criterion. In this Section, the complete derivation of the optimal solution for the AR case (OM-AR) is being presented for the first time. It is also shown that if the observation data is composed of true impulse response corrupted by Gaussianly distributed noise, then the proposed optimization produces the Maximum-Likelihood estimates (MLE) of the AR parameters. For other types of noise or deviations least-squares estimates (LSE) are found.

A critical step in the theoretical derivation of the error criterion is to decouple the multidimensional criterion into a non-linear problem for the AR-parameters and a linear problem for the numerator coefficient. In the decoupled form, the fitting error is found to be related to an equation error which is different than the ones that appear in CM or AM. But the form of the equation error is shown to be mathematically appropriate for the AR case. The non-linear criterion possesses inherent matrix prefiltering structure which directly leads to formulating an efficient iterative computational algorithm for its minimization. Several simulation examples demonstrate the superior performance of the proposed approach when compared to some of the existing suboptimal methods.

The Section is arranged as follows : in Subsection II, the problem is defined, the connection with MLE is established and some existing results are briefly outlined. In Subsection III the error criterion is theoretically derived for the AR case and the computational algorithm is presented. In Subsection IV, several simulation examples are given. Finally, in Subsection V, some concluding remarks are given.

II. Problem Statement and Previous Results

The z-domain transfer function for an auto-regressive model can be represented as,

$$H(z) = \frac{n_0}{1 + d_1 z^{-1} + \cdots + d_{p-1} z^{-(p-1)} + d_p z^{-p}} \triangleq \frac{n_0}{D(z)}, \tag{1}$$

where the coefficient of the z^0 term in the denominator has been assumed to be unity without any loss of generality. As an example, $H(z)$ may represent the transfer function of human vocal tract which is commonly modeled as an all-pole model. The model order p is assumed to be known. In case of speech signals, for example, a lot of experience and knowledge is already available and the value of $p = 10$ or 8 is usually chosen. An equivalent representation of the transfer function $H(z)$ can also be written in terms of its impulse response as,

$$H(z) \;=\; h(0) + h(1)z^{-1} + \cdots + h(N-2)z^{-(N-2)} + h(N-1)z^{-(N-1)} + \cdots . \tag{2}$$

The first N significant samples of $H(z)$ can be stacked in a vector form as,

$$\mathbf{h} \triangleq [h(0) \quad h(1) \quad \cdots \quad h(N-1)]^T . \tag{3}$$

Next, the vector containing the N samples of the 'prescribed' or 'desired' impulse response data is denoted as,

$$\mathbf{h}_P \triangleq \left[h_P(0) \quad h_P(1) \quad \cdots \quad h_P(N-1)\right]^T . \tag{4}$$

The desired IR data vector may represent the impulse response of vocal tract. With these definitions, the problem addressed in this Section may be stated as follows :

> Given a desired impulse response $\mathbf{h}_P$, the goal is to obtain the optimal estimates of the model parameters n_0 and $\mathbf{d}$ by minimizing the following least-squares IR model-fitting criterion :

$$\min_{n_0,\mathbf{d}} ||\mathbf{e}||^2 \triangleq \min_{n_0,\mathbf{d}} \sum_{i=0}^{N-1} \left[h_P(i) - \left\{\frac{n_0}{D(z)}\right\}\delta(i)\right]^2 . \qquad \text{where,} \tag{5}$$

$$\delta(i) = \begin{cases} 1, & i = 0 \\ 0, & i \neq 0, \end{cases} \tag{5a}$$

$$\mathbf{e} \triangleq \mathbf{h}_P - \mathbf{h} \qquad \text{and} \tag{5b}$$

$$\mathbf{d} \triangleq [1 \quad d_1 \quad \cdots \quad d_p]^T . \tag{5c}$$

The notation, $\left\{\frac{n_0}{D(z)}\right\}\delta(i)$ denotes the response of the system, $\frac{n_0}{D(z)}$ when driven by an input sequence, $\delta(i)$. Clearly, the criterion in (5) attempts to minimize the squared error between the desired and the estimated IR and hence, it can be expected to produce more accurate model than some well-known AR modeling methods (outlined below) which only minimize 'equation errors'. The least-squares problem in (5) is known to be nonlinear in $\mathbf{d}$ and standard nonlinear optimization algorithms have been utilized before in [15, 25-29, 36]. It should be emphasized that if the given IR-vector $\mathbf{h}_P$ is composed of the true IR-vector $\mathbf{h}$ and additive Gaussianly distributed noise or deviations then the minimization criterion in (5) is exactly equivalent to the maximization of the Likelihood criterion [see ref. 10, pp. 242-248]. Hence, for such a scenario the algorithm proposed in this Section produces the MLE of the AR-parameters. For all other types of noise and deviations the Least-Squares Estimates are found. It may also be noted that the MLE result in [10] is primarily based on the works in [5, 13, 44] where only the strictly proper ARMA case was considered. The MLE for transfer functions with arbitrary number of poles and zeros has been presented recently in [6].

In many applications, such as in linear prediction of speech signals [4, 8], only the estimation of the AR-parameters is of primary concern. The two most commonly used LP algorithms, AM and CM, do not solve the ideal problem stated in (5) whereas SMM attempts to solve the ideal problem by appropriate modification of the criterion in (5).. These three approaches are briefly summarized next.

Covariance Method [CM]

The ℓ_2-norm of the following equation error is minimized [7-10] :

$$\begin{bmatrix} h_P(p) & h_P(p-1) & \cdots & h_P(0) \\ \vdots & \vdots & \ddots & \vdots \\ h_P(N-1) & h_P(N-2) & \cdots & h_P(N-p-1) \end{bmatrix} \begin{bmatrix} 1 \\ d_1 \\ \vdots \\ d_P \end{bmatrix} = \mathbf{e}_{eq}^{CM}, \tag{6a}$$

$$\text{or,} \qquad \mathbf{H}_{CM}\mathbf{d} \triangleq \mathbf{e}_{eq}^{CM}. \tag{6b}$$

Note that $\mathbf{H}_{CM}$ is filled with available IR data only and hence $\mathbf{e}_{eq}^{CM}$ may be considered an 'exact' equation error.

Auto-correlation Method [AM]

In this case, the ℓ_2-norm of the following equation error is minimized [7-10] :

$$\begin{bmatrix} h_P(0) & 0 & 0 & \cdots & 0 \\ h_P(1) & h_P(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h_P(p) & h_P(p-1) & \cdots & \cdots & h_P(0) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h_P(N-1) & h_P(N-2) & \cdots & \cdots & h_P(N-p-1) \\ 0 & h_P(N-1) & \cdots & \cdots & h_P(N-p-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & h_P(N-1) \end{bmatrix} \begin{bmatrix} 1 \\ d_1 \\ \vdots \\ d_p \end{bmatrix} = \mathbf{e}_{eq}^{AM}, \tag{7a}$$

$$\text{or,} \qquad \mathbf{H}_{AM}\mathbf{d} \triangleq \mathbf{e}_{eq}^{AM}. \tag{7b}$$

The zeros in the upper and lower triangles of $\mathbf{H}_{AM}$ are not part of the prescribed IR-vector $\mathbf{h}_P$ and hence $\mathbf{e}_{eq}^{AM}$ is *not* an exact equation error.

It can be observed from (6) and (7) that the equation error for CM uses windowed data without making any prior assumptions about the data outside the observed window $\{h_P(0) \ \ldots \ h_P(N-1)\}$. On the other hand, AM uses unwindowed data but sets the data outside the observation frame to zero. Because of this reason, AM usually produces less accurate estimates than CM. But it should also be noted that even though $\mathbf{e}_{eq}^{AM}$ is not an exact equation error, one of the significant advantages of using AM is that the computationally efficient Levinson-Recursion algorithm can be utilized. In case of CM, a somewhat less efficient algorithm, Cholesky decomposition can be used [7-10]. Furthermore, the AR coefficients obtained by minimizing the norm of $\mathbf{e}_{eq}^{AM}$ produce a stable transfer function.

Steiglitz-McBride Method [SMM]

This method was originally developed for general ARMA parameter identification but it has also been adapted for AR parameter identification. For the AR case, the following modified fitting error criterion is optimized [7],

$$\min_{n_o,\mathbf{d}} \sum_{i=0}^{N-1} \left[\left\{ \frac{D(z)}{\hat{D}(z)} \right\} h_P(i) \ - \ \left\{ \frac{n_0}{\hat{D}(z)} \right\} \delta(i) \right]^2 . \tag{8}$$

The estimate $\hat{D}(z)$ obtained at any iteration step is used as a prefilter for obtaining the updated estimates at the succeeding iteration. Equation (8) closely approximates the criterion in (5) and both are identical if $D(z) = \hat{D}(z)$.

But the advantage of using (8) is that the unknown parameters in $\mathbf{d}$ and n_0 can be estimated by solving a set of simultaneous linear equations. It may be noted here that in [7], the numerator coefficient n_0 had been assumed to be unity but, in general, that may not be the case. The derivation of the proposed fitting error optimization scheme is in order.

III. Problem Formulation and Algorithm Development

In this Subsection, the multidimensional optimization problem in (5) is decoupled into a linear estimation problem for n_0 and a non-linear optimization problem for $\mathbf{d}$. Let $H_d(z)$ be the inverse filter corresponding to $D(z)$, *i.e.*,

$$D(z)H_d(z) \;=\; 1. \tag{9}$$

In time domain, this corresponds to a convolution operation where the d_k's are finite and the $h_d(n)$'s are infinite in extent. The first N significant terms of this convolution operation may be expressed in matrix notation as,

$$\mathbf{D}\mathbf{H}_d \;=\; \mathbf{I}_N \tag{10}$$

where, $\mathbf{I}_N$ denotes an $N \times N$ identity matrix,

$$\mathbf{D} \triangleq \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ d_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ d_p & d_{p-1} & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & d_p & \cdots & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times N} \qquad \text{and} \tag{11a}$$

$$\mathbf{H}_d \triangleq \begin{bmatrix} h_d(0) & 0 & \cdots & 0 \\ h_d(1) & h_d(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_d(N-1) & h_d(N-2) & \cdots & h_d(0) \end{bmatrix} \in \mathbb{R}^{N\times N}. \tag{11b}$$

Using (9), the expression in (1) can be rewritten as,

$$H(z) \;=\; \frac{n_0}{D(z)} \triangleq n_0 H_d(z). \tag{12}$$

Equating the first N coefficients of equal powers of z^{-1} in both sides of (12) and using vector notation,

$$\mathbf{h} \triangleq n_0\mathbf{h}_d, \tag{13}$$

where, $\mathbf{h}_d$ is also the first column of $\mathbf{H}_d$ defined in (11b), *i.e.*,

$$\mathbf{h}_d \triangleq [h_d(0)\ h_d(1)\ \cdots\ h_d(N-1)]^T. \tag{14}$$

With these definitions, the problem stated in (5) can be rephrased as,

$$\min_{n_0,\mathbf{d}} \|\mathrm{e}\|^2 \triangleq \min_{n_0,\mathbf{d}} \|\mathbf{h}_P \;-\; n_0\mathbf{h}_d\|^2, \tag{15}$$

where, the error vector is defined as,

$$\mathbf{e} \triangleq \mathbf{h}_P \;-\; n_0\mathbf{h}_d. \tag{16}$$

It is clear from (16) that the error $\mathbf{e}$ is linearly related to n_0 whereas $\mathbf{e}$ is non-linearly related to $\mathbf{d}$ through the vector $\mathbf{h}_d$. In this form, it is apparent that the present problem belongs to a class of mixed optimization problems where the linear and nonlinear variables appear separately. This class of problems has been studied extensively by numerical analysts [18-21]. In their work, the main objective had been to optimize the two sets of variables independently. Their argument goes as follows. If $\mathbf{h}_d$ (*i.e.*, $\mathbf{d}$) is known, then n_0 can be optimally estimated by the minimization of the criterion in (15) and the resulting least-squares estimate will be given by,

$$\hat{n}_0 \triangleq \mathbf{h}_d^{\#}\mathbf{h}_P, \tag{17}$$

where $^{\#}$ denotes pseudo-inverse operation defined as, $\mathbf{h}_d^{\#} \triangleq (\mathbf{h}_d^T\mathbf{h}_d)^{-1}\mathbf{h}_d^T$. In practice, $\mathbf{d}$ will not be known and it has to be estimated. Plugging $\hat{n}_0$ back in (15), the optimization criterion for $\mathbf{d}$ can be found as,

$$\min_{n_0,\mathbf{d}} ||\mathbf{h}_P - n_0\mathbf{h}_d||^2 \equiv \min_{\mathbf{d}} ||\mathbf{h}_P - (\mathbf{h}_d\mathbf{h}_d^{\#})\mathbf{h}_P||^2 \tag{18a}$$

$$= \min_{\mathbf{d}} ||(\mathbf{I}_N - \mathbf{P}_{\mathbf{h}_d})\mathbf{h}_P||^2, \tag{18b}$$

where $\mathbf{P}_{\mathbf{h}_d}$ denotes projection matrix defined as, $\mathbf{P}_{\mathbf{h}_d} \triangleq \mathbf{h}_d(\mathbf{h}_d^T\mathbf{h}_d)^{-1}\mathbf{h}_d^T$. For a larger class of multidimensional nonlinear optimization problems, it has been proved in Theorem 2.1 of [18] that if $\hat{\mathbf{d}}$ is estimated by minimizing the criterion in (18) and if that estimate is utilized for computing $\hat{n}_0$ using (17), then the resulting estimates are the ***unique and global minimizers*** of the criterion in (15). Hence, the original optimization problem in (5) is identical to the decoupled estimators in (17) and (18). This type of decoupled optimization of linear and non-linear subproblems had been utilized before in [5, 13, 45, 46] for strictly proper ARMA case and in [6] for the general ARMA case. The derivation for the AR case, as given here, appears to be new.

The AR-parameters in $\mathbf{d}$ are related to the error criterion in a complicated manner through $\mathbf{P}_{\mathbf{h}_d}$. Hence, the direct optimization of (18) *w.r.t.* $\mathbf{d}$ would require taking resort to standard non-linear optimization techniques such as Newton-Raphson or Gauss-Newton methods. Instead, following the strategy used in [6] for the general ARMA case, the criterion in (18) is reparameterized by relating it directly to the coefficients in $\mathbf{d}$. Appropriate partitioning of the matrices $\mathbf{D}$ and $\mathbf{H}_d$ gives,

$$\mathbf{D} \triangleq \left[\begin{array}{cccccc} 1 & 0 & \cdots & \cdots & 0 & 0 \\ \hline d_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ d_p & d_{p-1} & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & d_p & \cdots & \cdots & 1 \end{array}\right] \triangleq \left[\begin{array}{c} \mathbf{d}_u^T \\ \hline \mathbf{B}^T \end{array}\right] \qquad \text{and} \tag{19a}$$

$$\mathbf{H}_d \triangleq \left[\begin{array}{c|ccc} h_d(0) & 0 & \cdots & 0 \\ h_d(1) & h_d(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_d(N-1) & h_d(N-2) & \cdots & h_d(0) \end{array}\right] \triangleq \left[\mathbf{h}_d \ \vdots\ \mathbf{H}'_d\right]. \tag{19b}$$

Using these notations, the expression in (10) can be rewritten as,

$$\left[\begin{array}{c} \mathbf{d}_u^T \\ \hline \mathbf{B}^T \end{array}\right] \left[\mathbf{h}_d \ \vdots\ \mathbf{H}'_d\right] = \mathbf{I}_N, \tag{19c}$$

$$\text{or,} \quad \left[\begin{array}{c|c} \mathbf{d}_u^T\mathbf{h}_d & \mathbf{d}_u^T\mathbf{H}'_d \\ \hline \mathbf{B}^T\mathbf{h}_d & \mathbf{B}^T\mathbf{H}'_d \end{array}\right] = \left[\begin{array}{c|c} 1 & \mathbf{0}_{1\times(N-1)} \\ \hline \mathbf{0}_{(N-1)\times 1} & \mathbf{I}_{(N-1)} \end{array}\right]. \tag{19d}$$

The bottom-left corner element shows that the $N \times (N-1)$ matrix $\mathbf{B}$ and the vector $\mathbf{h}_d$ are orthogonal, *i.e.*, $\mathbf{B}^T\mathbf{h}_d = \mathbf{0}$. Also by construction,

$$rank(\mathbf{B}) + rank(\mathbf{h}_d) = N. \tag{20}$$

Hence, using a theorem on projection matrices [39],

$$\mathbf{P}_{\mathbf{B}} + \mathbf{P}_{\mathbf{h}_d} = \mathbf{I}_N. \tag{21}$$

Using this relationship in (18b), the following equivalent forms of reparameterized optimization criterion are obtained,

$$\begin{aligned}
\min_{\mathbf{d}} \|(\mathbf{I}_N - \mathbf{P}_{\mathbf{h}_d})\mathbf{h}_P\|^2 &= \min_{\mathbf{d}} \|\mathbf{P}_{\mathbf{B}}\mathbf{h}_P\|^2, && (22a)\\
&= \min_{\mathbf{d}} \|\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{h}_P\|^2, && (22b)\\
&= \min_{\mathbf{d}} \|\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{e}_{eq}\|^2, && (22c)\\
&= \min_{\mathbf{d}} \mathbf{e}_{eq}^T(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{e}_{eq}, && (22d)
\end{aligned}$$

where, $\mathbf{e}_{eq}$ is an equation error defined as,

$$\mathbf{e}_{eq} \triangleq \mathbf{B}^T\mathbf{h}_P. \tag{23}$$

This equation error can also be rewritten as,

$$\begin{aligned}
\mathbf{e}_{eq} = \mathbf{B}^T\mathbf{h}_P &= \begin{bmatrix} h_P(1) & h_P(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_P(N-1) & h_P(N-2) & \cdots & h_P(N-p-1) \end{bmatrix} \begin{bmatrix} 1 \\ d_1 \\ \vdots \\ d_p \end{bmatrix} && (24a)\\
&\triangleq \mathbf{H}_{AR}\mathbf{d}. && (24b)
\end{aligned}$$

A few observations may be made here regarding the equation error defined in (23). Clearly, $\mathbf{e}_{eq}$ differs from the equation errors used in CM and AM as defined in (6) and (7), respectively. The equation errors in those cases were formed in somewhat ad hoc manner on the basis of two types of autocorrelation estimates [7, 9, 10]. On the other hand, the particular form of equation error in (24a) resulted from purely mathematical consequences of the AR case under consideration. In particular, if the prescribed response $\mathbf{h}_P$ happens to be an exact impulse response of a p-th order AR transfer function, then the equation error in (24a) will be identically equal to zero, but the same will not be true for $\mathbf{e}_{eq}^{AM}$ in (7). The equation error for CM appearing in (6) will also be zero but $\mathbf{e}_{eq}^{CM}$ ignores the information contained in the upper $(p-1)$ equations of (24a). From this discussion, it can be concluded that more accurate estimates may be obtained if the equation error in (23) is used for the AR case. Minimization of this equation error will be utilized later in the computational algorithm for obtaining the initial estimate of $\mathbf{d}$.

Using (24b) in (22d), the reparameterized criterion can be expressed in the following useful form,

$$\min_{\mathbf{d}} \mathbf{d}^T\mathbf{H}_{AR}^T(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{H}_{AR}\mathbf{d}. \tag{25}$$

According to Theorem 2.1 of [18], the denominator vector $\mathbf{d}$ causing the minimum of the criterion in (25) is the desired optimum $\mathbf{d}^o$. The minimized error can then be found from,

$$\mathbf{e}^o = \mathbf{P}_{\mathbf{B}^o}\mathbf{h}_P, \tag{26}$$

where, $\mathbf{B}^o$ is constructed by using the optimum $\mathbf{d}^o$. The optimum estimate of the impulse response is then,

$$\mathbf{h}^o = \mathbf{h}_P - \mathbf{e}^o. \tag{27}$$

From [18], it can be also be inferred that if $\mathbf{h}_d^o$ is formed using $\mathbf{d}^o$ then the optimal n_0^o can be obtained using (17) as,

$$n_0^o \triangleq \mathbf{h}_d^{o\#}\mathbf{h}_P. \tag{28a}$$

Next it is shown that instead of using (28a), the optimal numerator coefficient can be found in a more straightforward manner as,

$$n_0^o = h^o(0), \tag{28b}$$

where $h^o(0)$ is the first sample of $\mathbf{h}^o$, the optimal impulse response estimate found in (27). In order to demonstrate this equivalence, equation (28b) is rewritten as,

$$n_0^o = \mathbf{h}_1^{oT}\mathbf{d}^o, \tag{29a}$$

where,

$$\mathbf{h}_1^o \triangleq [h^o(0)\ 0\ \cdots\ 0]^T. \tag{29b}$$

Note that the first term in $\mathbf{d}$ is always 1. Using the partitioning notation in (19a), equation (29a) can also be rewritten as,

$$\begin{aligned}
n_0^o &= \mathbf{d}_u^T\mathbf{h}^o, && && (30a)\\
&= \mathbf{d}_u^T(\mathbf{h}_P - \mathbf{e}^o), && \text{using (27)}, && (30b)\\
&= \mathbf{d}_u^T(\mathbf{h}_P - \mathbf{h}_P + \mathbf{h}_d^o\mathbf{h}_d^{o\#}\mathbf{h}_P), && \text{using (18a) and (26)} && (30c)\\
&= (\mathbf{d}_u^T\mathbf{h}^o{}_d)\mathbf{h}_d^{o\#}\mathbf{h}_P, && && (30d)\\
&= \mathbf{h}_d^{o\#}\mathbf{h}_P, && && (30e)
\end{aligned}$$

where, the last equality uses the fact that, $\mathbf{d}_u^T\mathbf{h}^o{}_d = 1$, which appears in the upper-left partition of (19c). This completes the proof of equivalence between the expressions in (28a) and (28b). It should be noted that (29) may be preferable over (28a) for computing n_0 because the computation of $\mathbf{h}_d^o$ and the pseudo-inverse solution required in (28a) can be avoided, whereas calculation of the optimal $\mathbf{h}^o$ in (27) may be a necessary step. Equations (22) and (29) are the two desired decoupled formulae for estimation of the coefficients of the denominator and numerator polynomials of the AR-model. It should be mentioned that unlike the decoupled forms of SMM given in [7] and [34], no approximations were introduced in deriving the decoupled estimators in (22) and (29). A computational algorithm for minimization of the criterion in (22) is outlined next.

Computational Algorithm

The criterion in (22) is non-linear in $\mathbf{d}$ and hence it can not be minimized directly. Standard gradient-based non-linear optimization techniques such as Newton-Raphson or Gauss-Newton algorithms could be used. But these algorithms utilize only the first few terms of Taylor series and are known to be highly sensitive to the choice of the initial estimates. But it can be observed from (25) that the error criterion possesses a good deal of matrix structure. Specifically, the expression appears to be a weighted quadratic criterion in $\mathbf{d}$, except that the weight matrix $(\mathbf{B}^T\mathbf{B})^{-1}$ itself is dependent on the unknowns in $\mathbf{d}$. This inherent mathematical structure of the criterion will be utilized to develop an iterative computational algorithm. The algorithm is similar to the ones for ARMA cases appearing in [5, 6, 13]. Here the complete derivation for the AR case will be given.

In order to initiate the iterative algorithm, $\mathbf{d}$ is first estimated by minimizing the ℓ_2-norm of the equation error $\mathbf{e}_{eq}$ defined in (24a). Partitioning $\mathbf{H}_{AR}$, the equation error $\mathbf{e}_{eq}$ can be rewritten as follows,

$$\mathbf{e}_{eq} = \mathbf{B}^T\mathbf{h}_P \triangleq \left[\begin{array}{c|cccc} h_P(1) & h_P(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_P(N-1) & h_P(N-2) & \cdots & h_P(N-p-1) \end{array}\right] \begin{bmatrix} 1 \\ d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix} \tag{31a}$$

$$\triangleq \left[\mathbf{g} \;\vdots\; \mathbf{G}\right] \mathbf{d}. \tag{31b}$$

Minimizing $||\mathbf{e}_{eq}||^2$ *w.r.t.* $\mathbf{d}$, the following initial estimate is obtained

$$\mathbf{d}^{(0)} = \begin{bmatrix} 1 \\ \cdots\cdots \\ -\mathbf{G}^{\#}\mathbf{g} \end{bmatrix}. \tag{32}$$

This estimate will be utilized for initiating the iterative computational algorithm. The final form of the error vector in (22a) is rewritten as,

$$\mathbf{e} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{h}_P \tag{33a}$$

$$\triangleq \mathbf{W}\mathbf{B}^T\mathbf{h}_P \qquad \text{using (24),} \tag{33b}$$

$$= \mathbf{W}\mathbf{H}_{AR}\mathbf{d} \qquad \text{using (31b),} \tag{33c}$$

$$= \mathbf{W}\left[\mathbf{g} \;\vdots\; \mathbf{G}\right]\mathbf{d} \tag{33d}$$

$$= \mathbf{W}\mathbf{g} + \mathbf{W}\mathbf{G}\hat{\mathbf{d}}, \tag{33e}$$

where,

$$\mathbf{W} \triangleq \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}, \qquad \text{and,} \tag{33f}$$

$$\hat{\mathbf{d}} \triangleq [d_1 \quad d_2 \quad \cdots \quad d_p]^T. \tag{33g}$$

If the matrix $\mathbf{W}$ is treated as independent of $\hat{\mathbf{d}}$, minimization of $||\mathbf{e}||^2$ *w.r.t.* $\hat{\mathbf{d}}$ results in the following estimate :

$$\begin{aligned} \hat{\mathbf{d}} &= -(\mathbf{W}\mathbf{G})^{\#}\mathbf{W}\mathbf{g} \\ &= -\left(\mathbf{G}^T\mathbf{W}^T\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{W}^T\mathbf{W}\mathbf{g}. \end{aligned} \tag{34}$$

But $\mathbf{W}$ does have dependence on the elements in $\hat{\mathbf{d}}$ and hence the estimate in (34) can only be computed iteratively. The estimate of $\mathbf{d}$ found in the i-th iteration step is used in (33f) to form $\mathbf{W}^{(i)}$ which is then utilized at the $(i+1)$-th step of iteration to compute $\mathbf{d}^{i+1}$ as follows :

$$\mathbf{d}^{(i+1)} = \begin{bmatrix} 1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ -[\mathbf{X}^{(i)}\mathbf{G}]^{-1}[\mathbf{X}^{(i)}]\mathbf{g} \end{bmatrix} \tag{35a}$$

where,

$$\mathbf{X}^{(i)} \triangleq \mathbf{G}^T\mathbf{W}^{T^{(i)}}\mathbf{W}^{(i)} \tag{35b}$$

$$= \mathbf{G}^T(\mathbf{B}^{T^{(i)}}\mathbf{B}^{(i)})^{-1}. \tag{35c}$$

The iterations are continued until $||\mathbf{d}_{i+1} - \mathbf{d}_i||^2 < \delta$, where δ is an arbitrarily small number.

It must be noted here that the iterations in (35) may not always converge to the absolute minimum of the error criterion in (5) and hence the estimated $\mathbf{d}$ may not be the optimum one. This is because in (35) the variability of $\mathbf{W}$ *w.r.t.* $\mathbf{d}$ has been ignored while minimizing $||\mathbf{e}||^2$. To achieve the optimum, the gradient of the complete expression of $||\mathbf{e}||^2$ must be set to zero. If desired, this may be done in a second phase of the algorithm which is outlined in the Appendix. Invoking Phase-2 will assure that at least a local minimum will be achieved. But it may be noted here that the simulation studies indicate that the Phase-1 of iterations using (35) does an excellent job of bringing the estimate very close to the optimum. It will be shown in Subsection IV that the Phase-2, if invoked, causes almost insignificant changes in the $\mathbf{d}$ vector and the minimized error norm. In simulations, the convergence was found to be quite rapid in both the phases. Once the estimates of $\mathbf{d}$ converge, n_0 is found using (26), (27) and (29), in sequence.

Discussion

The major computational burden of the algorithm is in performing the iterative refinement in (35), where, at each iteration step an $(N-1) \times (N-1)$ matrix $(\mathbf{B}^T\mathbf{B})$ needs to be inverted. But $(\mathbf{B}^T\mathbf{B})$ is a banded and symmetric matrix which can be inverted using computationally efficient Cholesky decomposition [8, 10]. Further reduction in computation is also possible because though $(\mathbf{B}^T\mathbf{B})$ is not purely Toeplitz, a major $(N-p) \times (N-p)$ diagonal block is symmetric-banded-Toeplitz and this block can be inverted with $O[(N-p)\log(N-p)] + O[p^2]$ operations [23]. The other $(p-1) \times (p-1)$ diagonal block is symmetric and can be inverted with $O[(p-1)^2]$ operations. Furthermore, the non-diagonal blocks contain mostly zero elements. Hence, using the block matrix inversion formula due to Schur [48], this matrix inversion can be computed with less than $O[(N-1)^2]$ operations. It may also be noted that in case of SMM the calculation of IR of the inverse filter and data filtering are required at every step of iteration, whereas the proposed method uses the estimated $\mathbf{d}$ directly to form the $\mathbf{B}$ matrix.

The LS error criterion defined in (5) attempts to match only the first N available samples of $h_P(n)$. No explicit assumption has been made about the unobserved samples, but the estimated rational transfer function essentially extends the impulse response beyond the observations. It may be noted here that minimum phase property can not be guaranteed with the AR-parameter estimates produced by the proposed algorithm. Extensive simulation studies indicate though that with converging IR sequences, the algorithms always produced stable solutions. It should also be pointed out that among existing methods, only the autocorrelation method can guarantee stable solutions. But AM uses windowed data and the IR fit with the estimates is usually not very accurate because the original least-squares IR error criterion is not minimized. To ensure minimum phase solution, AM can be used (instead of (32)) to obtain the initial estimates for starting the iterative AR-algorithm. If the estimates obtained from the iterative scheme becomes maximum phase at any iteration step of the AR-algorithm, the iterations can be terminated at that stage. The estimate found at the preceding iteration should be accepted as the best possible minimum phase solution that minimizes the optimal LS criterion in (5).

The model order selection problem has not been addressed in this work. It appears that for this essentially deterministic problem, Akaike Information Criterion (AIC) or Minimum Description Length Criterion (MDL) may not be applicable. But these criteria may be utilized when the prescribed impulse response data consists of true impulse response embedded in Gaussianly distributed noise.

The algorithm presented in this Section may also be quite useful for estimating MA filter coefficients. Presently, the most effective algorithm for MA modeling is Durbin's method [11] which, in fact, relies on two steps of AR parameter estimation. Traditionally, AM is utilized in both steps of Durbin's method because it produces minimum-phase polynomials [7, 9-11]. But the estimates obtained using AM may not be optimal because the true impulse response fitting error norm is not minimized. But the algorithm presented here produces *optimal* least-squares AR filter coefficients from prescribed impulse response data. Hence, it can be expected that the introduction of the proposed AR algorithm in one or both stages of Durbin's algorithm may produce more accurate MA parameter estimates.

It has been shown in [7] that the original SMM can also be decoupled into a linear and a non-linear subproblems. In a recent paper [34], the *strictly proper case* of the original SMM has been decoupled somewhat differently than in [7]. But more interestingly, it has also been demonstrated that the non-linear part of the decoupled SM criterion has exact mathematical equivalence with the optimal EFM criterion in [5]. It appears that using the new definitions of the matrices resulting from the matrix partitioning in (19), the AR-version of SMM as given in (8) can also be decoupled into linear and nonlinear subproblems. This equivalence may have an important consequence for the proposed algorithm. There already exists a convergence analysis of the *original* SM method in [47]. It can be hoped that the convergence analysis will also apply to the decoupled form of SM method given in [34]. If that happens to be the case, as alluded to in [34], the convergence analysis in [47] should also apply to the iterative computational algorithm presented in this work. It should be noted though that the results of SMM and the proposed optimal method may not be identical. Specifically, the numerator in the decoupled form of [34] is computed somewhat differently than (26) which is the optimal estimate. Furthermore, it should be also added that the iterative scheme in (35a) is not the only possible approach for iterative minimization of the equivalent criterion in (22). In fact, removing the requirement of $d_0 = 1$, the eigenvector corresponding to the minimum eigenvalue of the matrix $\mathbf{H}_{AR}^T(\mathbf{B}^{T^{(i)}}\mathbf{B}^{(i)})^{-1}\mathbf{H}_{AR}$ may also be used as $\mathbf{d}^{(i+1)}$, the estimate at the $(i+1)$-th iteration step [50]. This possibility is not obvious from the original SMM algorithm in [3].

IV. Simulation Results

In this Subsection, the performance of the proposed algorithm is evaluated by means of several AR(p) model identification examples with different p values. $\delta = 10^{-6}$ was used as the stopping criterion in both phases of the algorithm for all the examples below. The fitting-error norm defined in (5) was calculated at convergence using the estimated parameters and the results are tabulated in the 'Minimized Error Norm' column. Furthermore, in order to get a relative sense of performance, the logarithm of the ratios of the powers of the 'true IR' (known in these simulations) and the error powers are also tabulated in the 'Closeness in dB' columns.

Simulation 1 :

The desired impulse response has a Triangular form as shown by the solid lines in Fig. 1A - 1D. The resulting impulse response fit using Covariance method and Auto-correlation method are shown as connected circles in Figures 1A and 1B, respectively. The impulse response match at the end of each of the two phases of the algorithm described in Subsection III with $p = 4$ are shown in Fig. 1C and Fig. 1D, respectively. The minimized error norm and the closeness of the fit to the desired signal $\mathbf{h}_P$ are listed in Table 1. The number of iterations for convergence are also listed. It can be seen from the table and the figures that compared to AM and CM, the proposed scheme provided more accurate estimates. But it may also be observed that there is no significant difference in the results between the 1st and the 2nd phase of the proposed algorithm.

Table 1: Example 1: Comparison of three methods with Triangular Impulse Response

Method	Closeness in dB	Minimized Error Norm	Number of Iterations
Covariance	6.359	154.9	
Auto-Correlation	6.674	144.093	
Proposed Phase-1	23.436	3.037	5
Proposed Phase-2	23.439	3.035	3

Simulation 2 :

An arbitrary impulse response was generated with $p = 5$ for these simulations. If the algorithm in Subsection III is used directly to match the true response it will give perfect results. Instead, Gaussianly distributed white noise was added to the true response to obtain the desired response $\mathbf{h}_P$. Hence, the estimates obtained with the proposed algorithm will also be the MLE of the unknowns. For 20dB noise, the true and the desired responses are shown in Fig. 2A. The impulse response match using Covariance method and Autocorrelation method are shown in Figures 2B and 2C, respectively. The initial estimate obtained by minimizing the equation error in (24a) is shown in Fig. 2D. The impulse response fit obtained using the proposed algorithm at the end of Phase-1 and Phase-2 are shown in Fig. 2E and Fig. 2F, respectively. The minimized error norms and the closeness to the true response are listed in Table 2. It can be observed for this example that there is about 3dB difference in the impulse response fit between the two phases though the difference in the minimized error norms is quite small.

Table 2 : Example 2 : Comparison of three methods with 5-th order Impulse Response

Method	Closeness in dB	Minimized Error Norm	Number of Iterations
Covariance	1.027	1.847	
Auto-Correlation	12.442	0.691	
Equation Error in (24a)	15.459	0.656	
Proposed Phase-1	17.889	0.646	5
Proposed Phase-2	20.883	0.634	4

From these simulation results a fair conclusion may be drawn that the Phase-1 of the algorithm does an excellent job of error minimization. Hence, the Phase-2 of the algorithm need not be invoked for most applications. The results using SMM are close to the results at the end of Phase-1 if the original SMM [3] or the decoupled form in [34] are used. There were some numerical differences in the coefficients but the impulse response fit looked almost alike. The results with the AR-version of SMM given in [7] were poorer than Phase-1 results because the numerator coefficient is set to 1 in [7]. Extensive simulations with other examples show equivalent performance. Interestingly, the simulations also indicate that the proposed algorithm is quite immune to the choice of initial estimates. In fact, when CM or AM were used in place of (32) for obtaining the initial estimates, the results obtained at the end of Phase-1 or Phase-2 turned out to be exactly identical to the results listed in the Tables. But with Covariance method as initial estimate, the Phase-1 sometimes took one or two extra iterations to converge. This important observation indicates the robustness of the proposed algorithm to the choice of initial estimates.

V. Concluding Remarks

In this Section, a classical rational model identification problem has been addressed. The major focus was to develop an algorithm for optimal estimation of the parameters of an all-pole transfer function with arbitrary number of poles by model-fitting a prescribed impulse response. Unlike some existing results, no linearization or approximation has been done while deriving the theoretical optimization criterion. It is shown that the multidimensional non-linear problem can be decoupled into two smaller problems of which one is a linear problem and the other one is a non-linear problem. The inherent mathematical structure of the non-linear part is utilized to formulate an efficient iterative computational algorithm for estimating the denominator parameters. Global optimality properties of the estimators have been confirmed by relating the multidimensional optimization problem to certain well-known results in numerical analysis. In simulation studies also, the method has been shown to be highly effective. Regarding possible future work, it may be noted that most of the existing suboptimal 1-D algorithms have been extended for estimating 2-D filter coefficients from 2-D spatial domain data [22, 29, 35, 36, 40-43]. The possibility of formulating an *optimal* 2-D AR-filter design technique by extending the proposed method is being studied [49]. Extension of this work for identification of Multidimensional AR-systems from multidimensional impulse response data [38] is also under progress.

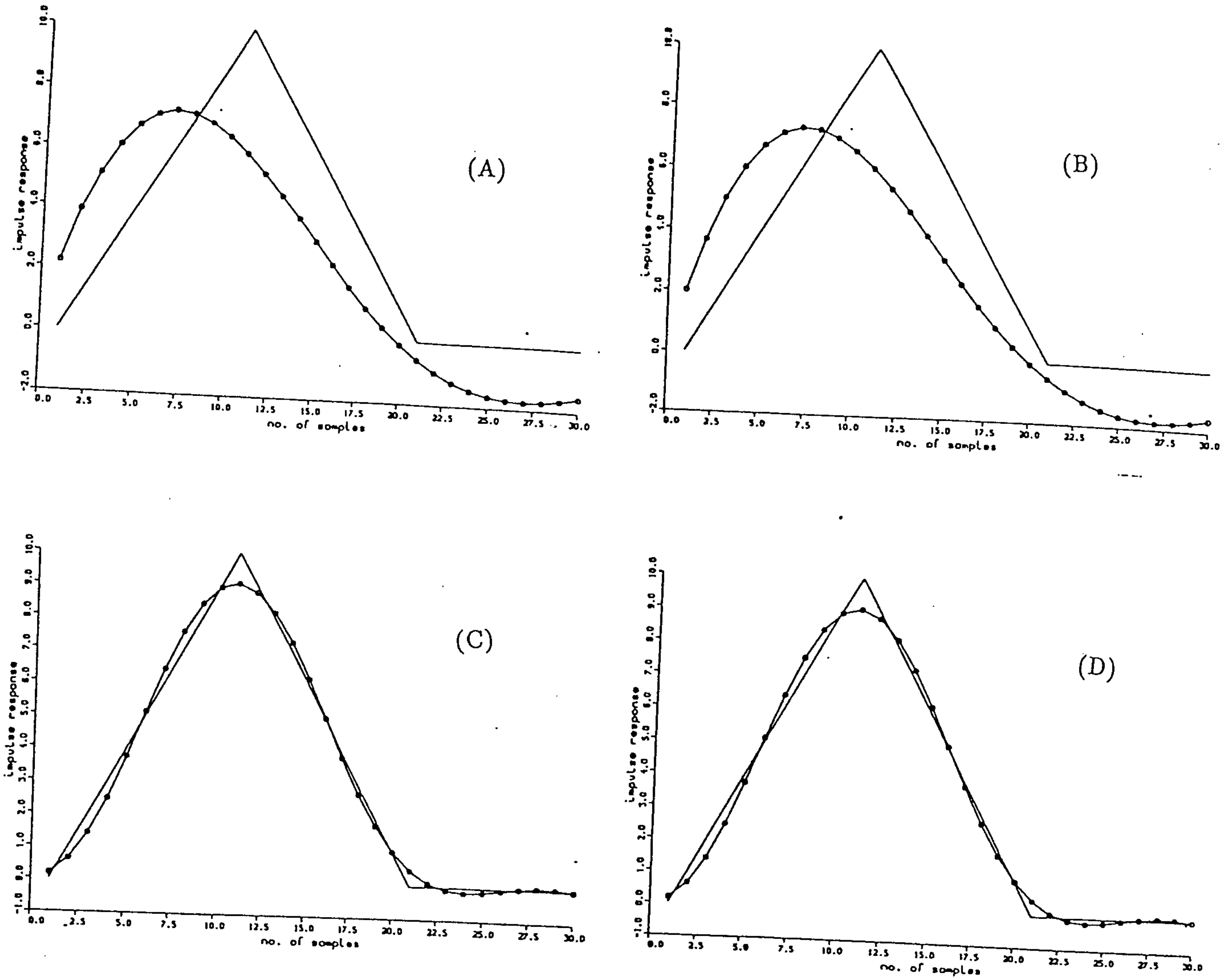


Fig. 1 : Simulation 1 :- A triangular impulse response is modeled by an AR(4) model. $\delta = 10^{-6}$ was used in both phases. The solid lines denote the prescribed impulse response and the connected circles show the fit with (A) Covariance Method, (B) Autocorrelation Method, (C) after Phase-1 convergence and (D) after Phase-2 convergence of the proposed method.

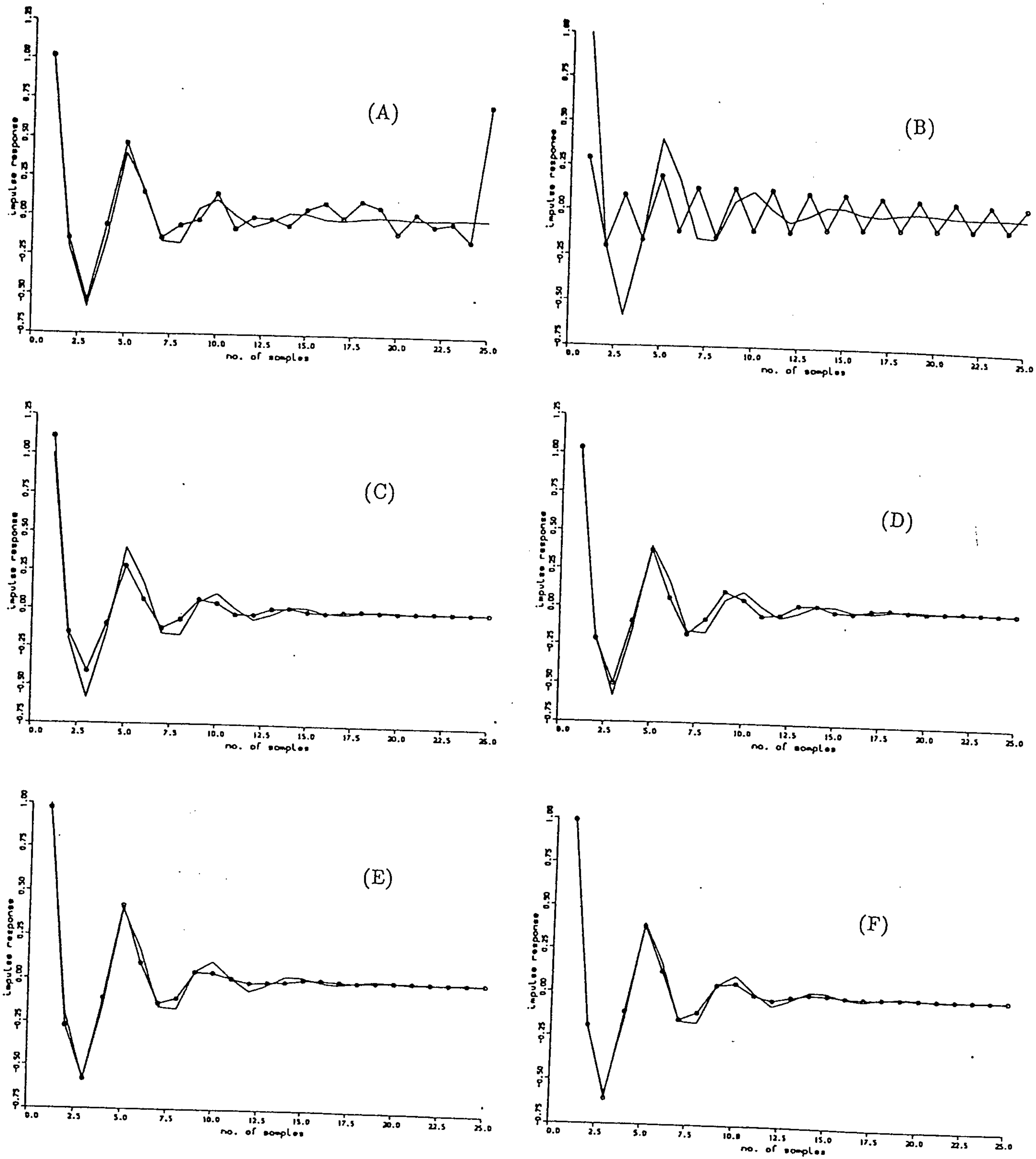


Fig. 2 : Simulation 2 :- 20dB noise was added to a true AR(5) impulse response to form the desired response (h_d). $\delta = 10^{-6}$ was used as stopping criterion. The solid lines denote the true impulse response. In Fig. 2A, the connected circles show the noisy signal h_d. In the other plots the connected circles show the fit with (B) Covariance Method, (C) Autocorrelation Method, (D) minimization of Equation error in (22f), (E) Phase-1 and (F) Phase-2 convergence.

REFERENCES

[1] R. E. Kalman, “Design of a Self Optimizing Control System,” *Trans. ASME*, Vol. 80, pp. 468-478, 1958.

[2] J.L.Shanks, “Recursion Filters for Digital Processing”, *Geophysics*, Vol. 32, pp. 33-51, 1967.

[3] K. Steiglitz and L.E. McBride, “A Technique for Identification of Linear Systems”, *IEEE Transactions on Automatic Control*, Vol. AC-10, pp. 461-464, 1965.

[4] K. Steiglitz, “On the Simultaneous Estimation of Poles and Zeros in Speech Analysis,” *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-25, no. 3, pp. 229-234, June, 1977.

[5] A.G. Evans and R. Fischl, “Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters”, *IEEE Transactions on Audio and Electro-Acoustics*, Vol. AU-21, pp. 61-65, 1973.

[6] A. K. Shaw, “Optimal Identification of Discrete-Time Systems From Impulse Response Data,” To be published, *IEEE Transactions on Acoustics, Speech and Signal Processing,* Jan., 1994.

[7] L.B. Jackson, *Digital Filters and Signal Processing*, Kluwer, Boston, 1986.

[8] J. Makhoul, “Linear Prediction : A Tutorial Review,” *Proceedings of the IEEE*, vol. 63, pp. 561-580, April, 1975.

[9] S.M. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[10] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[11] J. Durbin, “Efficient Estimation of Parameters in Moving-Average Models,” *Biometrika*, vol. 46, pp. 306-316, 1959.

[12] A. Stefanski and C. Weygandt, “Extension of the Steiglitz and McBride Identification Technique,” *IEEE Transactions on Automatic Control*, pp. 503-504, Oct., 1971.

[13] R. Kumaresan, L. L. Scharf and A. K. Shaw, “An Algorithm for Pole-Zero Modeling and Spectral Estimation,” *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1986.

[14] L. Ljung, *System Identification: Theory for the Users*, Prentice Hall, NJ, 1987.

[15] C. S. Burrus, T. W. Parks, and T. B. Watt, “A Digital Parameter-Identification Technique Applied to Biological Signals,” *IEEE Trans. on Bio-Medical Engineering*, vol. BME-18, pp. 35-37, Jan., 1971.

[16] Y. Bressler and A. Macovski, “Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 10, pp. 1081-1089, Oct., 1986.

[17] T. Söderström and P. Stoica, *System Identification*, Prentice Hall, NJ, 1987.

[18] G. H. Golub and V. Pereyra, “The Differentiation of Pseudoinverses and Nonlinear Problems Whose Variables Separate,” *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413-432, Apr., 1973.

[19] H. D. Scolnik, “On the Solution of Nonlinear Least Squares Problems,” *Proceedings of IFIP-1971*, Numerical Mathematics, North Holland, Amsterdam, pp. 18-23, 1971.

[20] M. R. Osborne, [1975], “Some Special Nonlinear Least Squares Problems,” *SIAM Journal on Numerical Analysis*, vol. 12, no. 4, pp. 571-592, Sep., 1975.

[21] I. Guttman, V. Pereyra, "Least Squares Estimation for a Class of Non-Linear Models," *Technometrics*, vol. 15, no. 2, pp. 209-218, May, 1973.

[22] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[23] A. K. Jain, "Fast Inversion of Banded Toeplitz Matrices by Circular Decompositions," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26, no. 2, pp. 121-126, April, 1978.

[24] E. R. Schulz, "Estimation of Pulse Transfer Function Parameters by Quasilinearization," *IEEE Transactions on Automatic Control*, pp. 424-426, 1968.

[25] K. Steiglitz, "Computer-Aided Design of Recursive Digital Filters," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-18, no. 2, pp. 123-129, June, 1970.

[26] A. G. Deczky, "Synthesis of Recursive Digital Filters Using the Minimum p-Error Criterion," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-20, no. 4, pp. 257-263, June, 1970.

[27] R. Fischl, "Optimal l_p-Approximation of Prescribed Impulse Response Functions on a Finite Point Set," in *Proc. IEEE Int. Symp. on Circuit Theory*, pp. 155-156, 1970.

[28] F. Brophy and A. C. Salazar, "Recursive Filter Synthesis in the Time Domain," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-22, no. 1, pp. 45-55, Feb., 1974.

[29] M. S. Bertran, "Approximation of Digital Filters in One and Two Dimensions," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-23, no. 5, pp. 438-443, Oct., 1975.

[30] C. S. Burrus and T. W. Parks, "Time Domain Design of Recursive Digital Filters," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-18, pp. 137-141, June, 1970.

[31] C. Charalambous, "Minimax Optimization of Recursive Digital Filters Using Recent Minimax Results," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-23, no. 4, pp. 333-345, June, 1977.

[32] M. J. Levin, "Estimation of a System Pulse Transfer Function in the Presence of Noise," *IEEE Transactions on Automatic Control*, pp. 229-235, 1964.

[33] G. Miller, "Least-Squares Rational Z-Transform Approximation," *Journal of the Franklin Institute,* vol. 295, no. 1, pp. 1-7, Jan., 1973.

[34] J. H. McClellan and D. Lee, "Exact Equivalence of the Steiglitz-McBride Iteration and IQML," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. ASSP-39, no. 2, pp. 509-512, Feb., 1991.

[35] J.L. Shanks, S. Treitel and J.H. Justice, "Stability and Synthesis of Two-Dimensional Recursive Filters" *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-20, pp. 115-128, 1972.

[36] J. A. Cadzow, "Recursive Digital Filter Synthesis via Gradient Based Algorithms", *IEEE Transaction on Acoustic, Speech and Signal Processing*, Vol. ASSP-24, pp. 349-355, 1976.

[37] A. K. Shaw and P. Misra, "Time Domain Identification of Proper Discrete Systems from Measured Impulse Response Data,", *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1689-1692, Toronto, Canada, May, 1991.

[38] A. K. Shaw, P. Misra and R. Kumaresan, "Identification of Multi-Dimensional Systems From Impulse Response Data," *30th IEEE Conf. on Dec. and Contr.*, Brighton, UK, Dec. 1991.

[39] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications.* New York, Wiley, 1971.

[40] D.E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Englewood Cliffs, NJ, Prentice Hall, 1984.

[41] G. A. Shaw and R. M. Mersereau, "Design, Stability and Performance of Two-Dimensional Recursive Digital Filters", *Tech. Report E21-B05-1*, Georgia Inst. of Technology School of Electrical Engg., 1979.

[42] T. Hinamoto and S. Maekawa, "Separable-Denominator 2-D Rational Approximation via 1-D Based Algorithm ", *IEEE Transactions on Circuits and Systems*, Vol. CAS-32, pp. 989-999, Nov., 1985.

[43] T. Hinamoto, "Design of 2-D Separable-Denominator Recursive Digital Filters", *IEEE Transactions on Circuits and Systems*, Vol. CAS-31, pp. 925-933, Nov., 1984.

[44] R. Kumaresan and A. K. Shaw, "High Resolution Bearing Estimation Without Eigendecomposition," *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, Florida, April, 1985.

[45] A.K. Shaw and R. Kumaresan, "Some Structured Matrix Approximation Problems", *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, New York, NY, pp. 2324-2327, April, 1988.

[46] R. Kumaresan and A.K. Shaw, "Superresolution by Structured Matrix Approximation", *IEEE Transactions on Antennas and Propagation*, Vol. AP-36, pp. 34-44, Jan., 1988.

[47] P. Stoica and T. Söderström, "The Steiglitz-McBride Identification Algorithm Revisited - Convergence Analysis and Accuracy Aspects," *IEEE Transactions on Automatic Control*, vol. AC-26, no. 3, pp. 712-717, June, 1981.

[48] E. Bodewig, *Matrix Calculus*, North-Holland Publishing Co., Amsterdam, 1956.

[49] A. K. Shaw, "Optimal Design of Two-Dimensional Filters from Spatial Domain Data," under progress.

[50] K. Hoffman and R. Kunze, *Linear Algebra*, Prentice Hall, Englewood Cliffs, New Jersey, 1971.

APPENDIX Computational Algorithm : Phase II

The second phase of the iterative algorithm is described in this Appendix. In this phase, the derivative of the matrix $\mathbf{W}$ *w.r.t.* $\hat{\mathbf{b}}$ is taken into consideration while minimizing the fitting error norm. The complete expression of the ℓ_2-norm of the error can be written as,

$$||\mathbf{e}||_2^2 \;=\; \mathbf{e}^T\mathbf{e} \;=\; (\mathbf{W}\mathbf{g} + \mathbf{W}\mathbf{G}\hat{\mathbf{d}})^T(\mathbf{W}\mathbf{g} + \mathbf{W}\mathbf{G}\hat{\mathbf{d}}). \qquad (A.1)$$

By setting the derivative of this squared norm to zero, the updated $\hat{\mathbf{b}}^{(i+1)}$ at the $(i+1)$-th iteration is given by,

$$\hat{\mathbf{b}}^{(i+1)} \;=\; -\,[\mathbf{U}^{(i)}\mathbf{G}]^{-1}[\mathbf{U}^{(i)}]\mathbf{g} \qquad (A.2)$$

where (suppressing the superscript $^{(i)}$),

$$\mathbf{U} \triangleq \mathbf{L}^T\mathbf{W} + \mathbf{G}^T\mathbf{W}^T\mathbf{W}, \qquad (A.2a)$$

$$\mathbf{L} \triangleq \left[\frac{\partial \mathbf{W}}{\partial d_1}\mathbf{e}_{eq} \;\middle|\; \cdots \middle| \frac{\partial \mathbf{W}}{\partial d_p}\mathbf{e}_{eq}\right], \qquad (A.2b)$$

$$\frac{\partial \mathbf{W}}{\partial d_k} \triangleq \frac{\partial}{\partial d_k}[\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}] = \frac{\partial \mathbf{B}}{\partial d_k}(\mathbf{B}^T\mathbf{B})^{-1} - \mathbf{W}\left[\left[\frac{\partial \mathbf{B}^T}{\partial d_k}\right]\mathbf{B} + \mathbf{B}^T\left[\frac{\partial \mathbf{B}}{\partial d_k}\right]\right](\mathbf{B}^T\mathbf{B})^{-1} \quad \text{and} \tag{A.2c}$$

$\frac{\partial \mathbf{B}}{\partial d_k}$ has the same form as the **B** matrix defined in (19a) but filled with all zeros except at the locations where d_k appear. For example,

$$\frac{\partial \mathbf{B}}{\partial d_p} = \begin{bmatrix} 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \cdots & 0 & 0 & \ddots & 1 \\ 0 & \cdots & 0 & 0 & \ddots & 0 \\ \vdots & \cdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{N \times N-1} \tag{A.2d}$$

Once $\hat{\mathbf{b}}^{(i+1)}$ is found, $\mathbf{b}^{(i+1)}$ can be formed as,

$$\mathbf{b}^{(i+1)} = \begin{bmatrix} 1 \\ \cdots\cdots \\ \hat{\mathbf{b}}^{(i+1)} \end{bmatrix} \tag{A.3a}$$

$$= \begin{bmatrix} 1 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ -[\mathbf{U}^{(i)}\mathbf{G}]^{-1}[\mathbf{U}^{(i)}]\mathbf{g} \end{bmatrix}. \tag{A.3b}$$

This minimization phase continues until $\mathbf{b}^{i+1} \simeq \mathbf{b}^i$ is reached and this optimum $\mathbf{b}^o$ vector corresponds to a minimum of the error surface of $||\mathbf{e}||_2^2$.

Section - 3.4 : DESIGN OF DENOMINATOR-SEPARABLE 2-D IIR FILTERS

SUMMARY

Optimal design of an important class of two-dimensional (2-D) digital IIR filters from spatial impulse response data is addressed. The denominator of the desired 2-D filter is assumed to be separable into two 1-D factors. The filter coefficients are estimated by minimizing the ℓ_2-norm of the error between the prescribed and the estimated spatial domain responses. The denominator and numerator estimation problems are theoretically decoupled into separate problems. The decoupled criteria have reduced dimensionality. The denominator criterion is simultaneously optimized *w.r.t.* the coefficients in both dimensions using an iterative algorithm. The numerator coefficients are found in a straight-forward manner. If the desired response is known to be symmetric, the proposed algorithm can be constrained to produce optimal denominators which are identical in both domains. The performance of the algorithm is demonstrated with simulation examples.

I. Introduction

Two-Dimensional IIR filters are commonly used in image processing and 2-D filtering. Synthesis of such filters from prescribed spatial domain impulse response data is an important and challenging design problem and has received considerable attention in recent literature [1, 2, 4, 5, 8, 12, 14, 16]. Spatial-domain design of 2-D IIR filters is analogous to 1-D recursive filter design based on time-domain specifications. Most 2-D filter design algorithms are basically extensions of existing 1-D algorithms. In particular, Shanks *et al* [12] had extended the work of Shanks [11]; Cadzow [1] and Shaw and Mersereau [16] utilized many of the general non-linear optimization methods; and Shaw and Mersereau [16] also extended the work of Steiglitz and McBride [17]. The 1-D work of Mullis and Roberts [7] was further extended and applied to the 2-D case in [5].

The approaches noted above do not minimize the true spatial impulse response error, though it may be mentioned that the extension of Steiglitz-McBride method in [16] closely approximates the true fitting error. For the *strictly-proper* case, *i.e.*, when the numerator order is one less than that of the denominator, Evans and Fischl (EFM) had proposed an optimal method for synthesis of 1-D IIR filters [3]. The 2-D filter synthesis algorithm presented here is a generalization of EFM to 2-D. Proposed solution is optimal in the sense that it minimizes true and complete spatial error criterion for the design of strictly-proper 2-D IIR filters.

EFM has been found to be highly accurate for 1-D filter design. A modified complex version of the EFM with certain symmetry constraints has also been shown to be effective for maximum-likelihood 1-D and 2-D frequency-wavenumber estimation [8, 13, 15]. Generalization of EFM for strictly-proper 2-D filter design has also been considered previously [4, 5], but it appears that the full potential of EFM has not been utilized in the 2-D case. Specifically, it will be shown that the complete error criterion encompassing the entire subspace orthogonal to the model fitting error was not optimized in [4, 5]. Instead, two suboptimal error criteria were formed in each domain and the filter coefficients were optimized in the two dimensions independently.

In this Section, a 2-D version of EFM is developed for *optimal* design of 2-D recursive filters from prescribed spatial domain data. The complete basis space orthogonal to the spatial fitting error will be identified and the corresponding error criterion will be shown to be dependent only on the 2-D filter parameters. Similar to 1-D EFM, the non-linear error criterion will be decoupled into a purely linear and a non-linear sub-problem. For the separable denominator case, it is also shown that the error vector possesses a *quasi-linear* relationship with the denominator coefficients in both domains simultaneously. Unlike several existing 2-D methods [1, 3, 4, 5], the *exact* fitting error is minimized *w.r.t.* the filter coefficients in both dimensions simultaneously. Simultaneous optimization is particularly effective for synthesizing 2-D filters with symmetric impulse response which are quite common in practice. In such cases, the criterion can be constrained to produce identical denominators in both domains ensuring symmetry in the estimated spatial response.

The Section is arranged as follows : In Subsection II, the least-squares problem is stated. In Subsection III, the preliminaries for the non-separable numerator and separable denominator case is given. In Subsection IV, the new orthogonal basis spaces are defined, the error criterion is derived and the computational algorithm is summarized. In Subsection V, some simulation results are given and finally, in Subsection VI some concluding remarks along with directions for future work are included.

II. Problem Statement and Formulation

In general, a 2-D rational function $H(z_1, z_2)$, with non-decomposable numerator and denominator is described as :

$$H(z_1,z_2) = \frac{Q(z_1,z_2)}{P(z_1,z_2)} = \frac{\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} q(i,j) z_1^{-i} z_2^{-j}}{\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} p(i,j) z_1^{-i} z_2^{-j}}. \tag{1}$$

Note that for the strictly-proper case of EFM, $n_1 = m_1 - 1$ and $n_2 = m_2 - 1$. If the $k_1 \times k_2$ first quadrant samples are assumed to be significant, $H(z_1, z_2)$ can also be written as,

$$H(z_1,z_2) = \mathbf{z}_1^T \mathbf{H} \mathbf{z}_2 \tag{2}$$

where, $\mathbf{z}_1 \triangleq [1 \ z_1^{-1} \cdots z_1^{-(k_1-1)}]^T$, $\mathbf{z}_2 \triangleq [1 \ z_2^{-1} \cdots z_2^{-(k_2-1)}]^T$ and

$$\mathbf{H} \triangleq \begin{bmatrix} h(0,0) & h(0,1) & \cdots & h(0,k_2-1) \\ h(1,0) & h(1,1) & \cdots & h(1,k_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(k_1-1,0) & h(k_1-1,1) & \cdots & h(k_1-1,k_2-1) \end{bmatrix}. \tag{3}$$

Define a vector by stacking the columns $\mathbf{H}$ as follows :

$$\mathbf{h} \triangleq \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{k_2} \end{bmatrix} \tag{4}$$

where, $\mathbf{h}_i$ denotes the i^{th} column of $\mathbf{H}$. Next, let the *prescribed* space-domain impulse response matrix be denoted as,

$$\mathbf{X} \triangleq \begin{bmatrix} x(0,0) & x(0,1) & \cdots & x(0,k_2-1) \\ x(1,0) & x(1,1) & \cdots & x(1,k_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k_1-1,0) & x(k_1-1,1) & \cdots & x(k_1-1,k_2-1) \end{bmatrix} \tag{5a}$$

$$\triangleq [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_{k_2}] \tag{5b}$$

and the corresponding vector be formed as,

$$\mathbf{x} \triangleq \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{k_2} \end{bmatrix}. \tag{6}$$

In this Section, the following 2-D least-squares synthesis problem is addressed :

Given the 2-D spatial impulse response matrix $\mathbf{X}$, estimate $\mathbf{p}$ and $\mathbf{q}$ by optimizing the following error criterion,

$$\min_{\mathbf{q},\mathbf{p}} ||\mathbf{e}||^2 \triangleq ||\mathbf{x} - \mathbf{h}||^2 \quad \text{with, } p(0,0) = 1, \quad \text{where,} \tag{7a}$$

$$\mathbf{q} \triangleq [q(0,0)\ q(0,1)\ \cdots\ q(n_1,n_2)]^T \qquad \text{and} \tag{7b}$$

$$\mathbf{p} \triangleq [p(0,0)\ p(0,1)\ \cdots\ p(m_1,m_2)]^T \tag{7c}$$

This problem is nonlinear in $\mathbf{p}$ and standard gradient-based optimization algorithms have been used for 1-D as well as for 2-D designs [1, 2, 16]. But these generic algorithms do not make effective use of the matrix-structures inherent in this particular problem and they are known to be sensitive to initialization. Several sub-optimal algorithms based on linearization of the true non-linear criterion have also been proposed [2, 11, 16, 17]. In this work, the exact fitting criterion defined in (7) will be theoretically *decoupled* into a purely linear problem for $\mathbf{q}$ and a non-linear problem for $\mathbf{p}$. Furthermore, the non-linear criterion will be shown to possess a quasi-linear relationship to the unknown denominator coefficients. This will lead to the formulation of an iterative algorithm for its minimization.

III. Design With Separable Denominator and Non-Separable Numerator

In this case, the 2-D rational transfer function can be written as,

$$H(z_1,z_2) = \frac{\sum_{i=0}^{m_1-1}\sum_{j=0}^{m_2-1} q(i,j) z_1^{-i} z_2^{-j}}{\sum_{i=0}^{m_1} c(i) z_1^{-i} \sum_{j=0}^{m_2} d(j) z_2^{-j}}. \tag{8a}$$

Define,

$$\mathbf{c} \triangleq [c(0)\ c(1)\ \ldots c(m_1)]^T \qquad \text{and} \tag{8b}$$

$$\mathbf{d} \triangleq [d(0)\ d(1)\ \ldots d(m_2)]^T. \tag{8c}$$

Multiplying both sides of (8a) by $\sum_{i=0}^{m_1} c(i) z_1^{-i} \sum_{j=0}^{m_2} d(j) z_2^{-j}$ and equating the coefficients of the same powers of z^{-1} [5],

$$[\mathbf{D}_1^T \otimes \mathbf{C}_1^T]\mathbf{h} = \mathbf{q} \tag{9a}$$

$$[\mathbf{D}_1^T \otimes \mathbf{C}^T]\mathbf{h} = \mathbf{0} \tag{9b}$$

$$[\mathbf{D}^T \otimes \mathbf{C}_1^T]\mathbf{h} = \mathbf{0} \tag{9c}$$

$$[\mathbf{D}^T \otimes \mathbf{C}^T]\mathbf{h} = \mathbf{0} \tag{9d}$$

$$[\mathbf{I}_{k_2} \otimes \mathbf{C}^T]\mathbf{h} = \mathbf{0} \tag{9e}$$

$$[\mathbf{D}^T \otimes \mathbf{I}_{k_1}]\mathbf{h} = \mathbf{0} \qquad \textbf{where,} \tag{9f}$$

$$\mathbf{C} \triangleq \begin{bmatrix} c(m_1) & 0 & \cdots & 0 \\ c(m_1-1) & c(m_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c(0) & c(1) & \cdots & c(m_1) \\ 0 & c(0) & \cdots & c(m_1-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c(0) \end{bmatrix} \in \mathbb{R}^{k_1\times(k_1-m_1)}, \tag{10a}$$

$$\mathbf{D} \stackrel{\triangle}{=} \begin{bmatrix} d(m_2) & 0 & \cdots & 0 \\ d(m_2-1) & d(m_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d(0) & d(1) & \cdots & d(m_2) \\ 0 & d(0) & \cdots & d(m_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d(0) \end{bmatrix} \in \mathbb{R}^{k_2 \times (k_2-m_2)}, \tag{10b}$$

$$\mathbf{C}_1 \stackrel{\triangle}{=} \begin{bmatrix} c(0) & c(1) & \cdots & c(m_1-1) \\ 0 & c(0) & \cdots & c(m_1-2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c(0) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{k_1 \times m_1}, \tag{10c}$$

$$\mathbf{D}_1 \stackrel{\triangle}{=} \begin{bmatrix} d(0) & d(1) & \cdots & d(m_2-1) \\ 0 & d(0) & \cdots & d(m_2-2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d(0) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{k_2 \times m_2}, \tag{10d}$$

where, $\mathbf{I}_{k_1} \in \mathbb{R}^{k_1 \times k_1}$ and $\mathbf{I}_{k_2} \in \mathbb{R}^{k_2 \times k_2}$ are identity matrices and $\otimes$ denotes the Kronecker product [9]. If $\mathbf{h}$, $\mathbf{c}$ and $\mathbf{d}$ are known, the numerator vector $\mathbf{q}$ can be calculated using (9a). But in practice, $\mathbf{h}$, $\mathbf{c}$ and $\mathbf{d}$ need to be estimated from the prescribed response $\mathbf{x}$. If $\mathbf{h}$ is replaced by the prescribed $\mathbf{x}$ in (9b)-(9f), the right hand sides will not be equal to zero. Instead, it will result in the following *equation errors* :

$$[\mathbf{D}_1^T \otimes \mathbf{C}^T]\mathbf{x} = [\mathbf{D}_1^T \otimes \mathbf{C}^T][\mathbf{h}+\mathbf{e}] = [\mathbf{D}_1^T \otimes \mathbf{C}^T]\mathbf{e} \stackrel{\triangle}{=} \mathbf{e}_{eq}^1 \tag{11a}$$

$$[\mathbf{D}^T \otimes \mathbf{C}_1^T]\mathbf{x} = [\mathbf{D}^T \otimes \mathbf{C}_1^T][\mathbf{h}+\mathbf{e}] = [\mathbf{D}^T \otimes \mathbf{C}_1^T]\mathbf{e} \stackrel{\triangle}{=} \mathbf{e}_{eq}^2 \tag{11b}$$

$$[\mathbf{D}^T \otimes \mathbf{C}^T]\mathbf{x} = [\mathbf{D}^T \otimes \mathbf{C}^T][\mathbf{h}+\mathbf{e}] = [\mathbf{D}^T \otimes \mathbf{C}^T]\mathbf{e} \stackrel{\triangle}{=} \mathbf{e}_{eq}^3 \tag{11c}$$

$$[\mathbf{I}_{k_2} \otimes \mathbf{C}^T]\mathbf{x} = [\mathbf{I}_{k_2} \otimes \mathbf{C}^T][\mathbf{h}+\mathbf{e}] = [\mathbf{I}_{k_2} \otimes \mathbf{C}^T]\mathbf{e} \stackrel{\triangle}{=} \mathbf{e}_{eq}^4 \tag{11d}$$

$$[\mathbf{D}^T \otimes I_{k_1}]\mathbf{x} = [\mathbf{D}^T \otimes I_{k_1}][\mathbf{h}+\mathbf{e}] = [\mathbf{D}^T \otimes I_{k_1}]\mathbf{e} \stackrel{\triangle}{=} \mathbf{e}_{eq}^5. \tag{11e}$$

In (11a)-(11e), the fact that $\mathbf{x} = \mathbf{h} + \mathbf{e}$ and the orthogonal relationships in (9b)-(9f) have been utilized. These equations show the relationships between the fitting error $\mathbf{e}$ and equation errors. As in case of 1-D EFM [7], in order to minimize $||\mathbf{e}||^2$ an inverse relationship of the form,

$$\mathbf{e} = \mathbf{W}(\mathbf{c}, \mathbf{d})\mathbf{e}_{eq} \tag{12}$$

need to be found. The matrix, $\mathbf{W}(\mathbf{c}, \mathbf{d})$ needs to be constructed using $\mathbf{c}$ and $\mathbf{d}$.

The problem of determining the denominator coefficient vectors $\mathbf{c}$ and $\mathbf{d}$ is essentially equivalent to the search for $(k_1k_2 - m_1m_2)$ linearly independent vectors orthogonal to $\mathbf{h}$. These orthogonal basis space must be dependent on the elements in $\mathbf{c}$ and $\mathbf{d}$ only. Equation (9a) clearly shows that $\mathbf{D}_1 \otimes \mathbf{C}_1 \in k_1k_2 \times m_1m_2$ can not be orthogonal to $\mathbf{h}$. On the other hand, (9b)-(9f) demonstrate that the matrices $\mathbf{D}_1 \otimes \mathbf{C}$, $\mathbf{D} \otimes \mathbf{C}_1$, $\mathbf{D} \otimes \mathbf{C}$, $\mathbf{I}_{k_2} \otimes \mathbf{C}$ and $\mathbf{D}^T \otimes I_{k_1}$ are indeed all orthogonal to $\mathbf{h}$. But summing the respective number of columns of these five full-rank matrices, the total turns out to be $(3k_1k_2 - m_1m_2 - k_1m_2 - k_2m_1)$ vectors of length k_1k_2 each. Hence, these set of matrices can not all be linearly independent of each other. In [5], the matrices, $\mathbf{D}_1 \otimes \mathbf{C}$, $\mathbf{D} \otimes \mathbf{C}_1$, $\mathbf{D} \otimes \mathbf{C}$, as formed in (9b)-(9d), respectively, were utilized to form an inverse relationship as required by (12). These set of matrices do contain $(k_1k_2 - m_1m_2)$ linearly independent vectors of length k_1k_2, but unfortunately, they are not orthogonal to each other.

It may be noted here that, in a previous 2-D generalization of EFM [4], the complete spatial fitting error criterion was not formulated. In a later work, it was partially formulated in equation (14) of [5], but no algorithm was presented for minimizing that criterion with respect to the unknown parameters. Instead, in both those works, two separate criteria were minimized independently. Specifically, (9e) and (9f) were used in [4, 5] to estimate $\mathbf{c}$ and $\mathbf{d}$ using two independent optimizations. But $\mathbf{I}_{k_2} \otimes \mathbf{C}$ and $\mathbf{D} \otimes I_{k_1}$ contain $(k_1k_2 - k_2m_1)$ and $(k_1k_2 - m_2k_1)$ linearly independent vectors, respectively. The entire $(k_1k_2 - m_1m_2)$ dimensional vector-space orthogonal to $\mathbf{h}$ was not optimized simultaneously *w.r.t* $\mathbf{c}$ and $\mathbf{d}$. It is not apparent if the optima of these separate criteria are identical to those of the true 2-D criterion. In Subsection IV, a new set of orthogonal vectors will be constructed which will lead to the formulation of an exact 2-D spatial fitting error criterion that can be optimized simultaneously *w.r.t.* $\mathbf{c}$ and $\mathbf{d}$.

IV. Formulation of the Orthogonal Basis Space :

According to orthogonality principle [15], the fitting error $\mathbf{e}$, at minimum, ought to be orthogonal to the 'estimated' $\mathbf{h}$ that minimizes the error. It is also desirable to have the resulting error criterion dependent on the denominator coefficients only. To meet these requirements two Vandermonde matrices are formed as follows,

$$\mathbf{T} \triangleq \begin{bmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_{m_1} \\ t_1^2 & \dots & t_{m_1}^2 \\ \vdots & \ddots & \vdots \\ t_1^{k_1-1} & \dots & t_{m_1}^{k_1-1} \end{bmatrix} \text{ and } \mathbf{Q} \triangleq \begin{bmatrix} 1 & \dots & 1 \\ q_1 & \dots & q_{m_2} \\ q_1^2 & \dots & q_{m_2}^2 \\ \vdots & \ddots & \vdots \\ q_1^{k_2-1} & \dots & q_{m_2}^{k_2-1} \end{bmatrix}, \tag{13}$$

where, $t_i = e^{j\omega_i}; \omega_i, i = 1, \dots, m_1$ and $q_i = e^{j\theta_i}; \theta_i, i = 1, \dots, m_2$ be the roots of the polynomials $C(z_1) = \sum_{i=0}^{m_1} c(i) z_1^{-i}$ and $D(z_2) = \sum_{i=0}^{m_2} d(i) z_2^{-i}$, respectively. Hence, by construction,

$$\mathbf{C}^T\mathbf{T} = \mathbf{0} \qquad \text{and} \tag{14a}$$

$$\mathbf{D}^T\mathbf{Q} = \mathbf{0} \tag{14b}$$

Furthermore, using (9e) and (9f),

$$[\mathbf{Q}^T \otimes \mathbf{C}^T]\mathbf{h} = [\mathbf{Q}^T \otimes \mathbf{I}][\mathbf{I} \otimes \mathbf{C}^T]\mathbf{h} = \mathbf{0} \qquad \text{and} \tag{15a}$$

$$[\mathbf{D}^T \otimes \mathbf{T}^T]\mathbf{h} = [\mathbf{I} \otimes \mathbf{T}^T][\mathbf{D}^T \otimes \mathbf{I}]\mathbf{h} = \mathbf{0}. \tag{15b}$$

The orthogonality relationships in (9d), (15a) and (15b) demonstrate that the three matrices $\mathbf{Q} \otimes \mathbf{C}$, $\mathbf{D} \otimes \mathbf{T}$ and $\mathbf{D} \otimes \mathbf{C}$ together constitute $(k_1k_2 - m_1m_2)$ dimensional vector space orthogonal to $\mathbf{h}$. Interestingly, these matrices are not only formed with linearly independent columns they are also *mutually orthogonal* to each other. This orthogonality claim can be easily substantiated as follows :

$$[\mathbf{Q}^T \otimes \mathbf{C}^T][\mathbf{D} \otimes \mathbf{T}] = [\mathbf{Q}^T\mathbf{D} \otimes \mathbf{C}^T\mathbf{T}] = [\mathbf{0} \otimes \mathbf{0}] = \mathbf{0}, \tag{16a}$$

$$[\mathbf{Q}^T \otimes \mathbf{C}^T][\mathbf{D} \otimes \mathbf{C}] = [\mathbf{Q}^T\mathbf{D} \otimes \mathbf{C}^T\mathbf{C}] = [\mathbf{0} \otimes \mathbf{C}^T\mathbf{C}] = \mathbf{0} \qquad \text{and} \tag{16b}$$

$$[\mathbf{D}^T \otimes \mathbf{T}^T][\mathbf{D} \otimes \mathbf{C}] = [\mathbf{D}^T\mathbf{D} \otimes \mathbf{T}^T\mathbf{C}] = [\mathbf{D}^T\mathbf{D} \otimes \mathbf{0}] = \mathbf{0}. \tag{16c}$$

It should also be mentioned here that the matrices $\mathbf{T}$ and $\mathbf{Q}$ are useful only in this intermediate stage of deriving the 2-D error criterion and they will not be needed in the final optimization steps.

If the vector $\mathbf{h}$ in (15) is replaced by $\mathbf{x}$, then similar to (11), the following equation errors are formed :

$$\begin{pmatrix} \mathbf{Q}^T \otimes \mathbf{C}^T \\ \mathbf{D}^T \otimes \mathbf{T}^T \\ \mathbf{D}^T \otimes \mathbf{C}^T \end{pmatrix} \mathbf{x} \;=\; \begin{pmatrix} \mathbf{Q}^T \otimes \mathbf{C}^T \\ \mathbf{D}^T \otimes \mathbf{T}^T \\ \mathbf{D}^T \otimes \mathbf{C}^T \end{pmatrix} \mathbf{e} \triangleq \mathbf{e}_{eq}. \tag{17}$$

The optimized $\mathbf{e}$ must be orthogonal to the estimated $\mathbf{h}$, whereas the three matrices, $\mathbf{Q}\otimes\mathbf{C}$, $\mathbf{D}\otimes\mathbf{T}$ and $\mathbf{D}\otimes\mathbf{C}$, were shown to be orthogonal to $\mathbf{h}$ in (9d), (15a) and (15b), respectively. Hence, $\mathbf{e}$ can be constructed as a linear combination of the columns of these orthogonal set of matrices, *i.e.*,

$$\mathbf{e} \;=\; (\,\mathbf{Q}\otimes\mathbf{C} \quad \mathbf{D}\otimes\mathbf{T} \quad \mathbf{D}\otimes\mathbf{C}\,)\mathbf{f} \tag{18}$$

where,

$$\mathbf{f} \triangleq [f_1 \;\; f_2 \;\; \cdots \;\; f_{k_1k_2-m_1m_2}]^T \tag{19}$$

is a vector of constants which are to be determined. Using this form of $\mathbf{e}$ in (17), the equation error can be written as,

$$\begin{pmatrix} \mathbf{Q}^T \otimes \mathbf{C}^T \\ \mathbf{D}^T \otimes \mathbf{T}^T \\ \mathbf{D}^T \otimes \mathbf{C}^T \end{pmatrix} \mathbf{x} \;=\; \begin{pmatrix} \mathbf{Q}^T\mathbf{Q} \otimes \mathbf{C}^T\mathbf{C} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^T\mathbf{D} \otimes \mathbf{T}^T\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}^T\mathbf{D} \otimes \mathbf{C}^T\mathbf{C} \end{pmatrix} \mathbf{f} \triangleq \mathbf{e}_{eq}. \tag{20}$$

The matrix on the r.h.s. is square block-diagonal with square diagonal blocks and hence it can be inverted to uniquely determine the vector of constants $\mathbf{f}$ as,

$$\begin{aligned} \mathbf{f} \;&=\; \begin{pmatrix} (\mathbf{Q}^T\mathbf{Q})^{-1} \otimes (\mathbf{C}^T\mathbf{C})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{D}^T\mathbf{D})^{-1} \otimes (\mathbf{T}^T\mathbf{T})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{D}^T\mathbf{D})^{-1} \otimes (\mathbf{C}^T\mathbf{C})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Q}^T \otimes \mathbf{C}^T \\ \mathbf{D}^T \otimes \mathbf{T}^T \\ \mathbf{D}^T \otimes \mathbf{C}^T \end{pmatrix} \mathbf{x} && (21a) \\ &=\; \begin{pmatrix} (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T \otimes (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \\ (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T \otimes (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T \\ (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T \otimes (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{D}^T \end{pmatrix} \mathbf{x}. && (21b) \end{aligned}$$

Using this expression of $\mathbf{f}$ in (18) the fitting error becomes,

$$\mathbf{e} \;=\; [\mathcal{P}_{\mathbf{Q}} \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{T}} + \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}}]\mathbf{x} \tag{22}$$

where, $\mathcal{P}_{[\cdot]}$ denotes the Projection matrix, *e.g.*, $\mathcal{P}_{\mathbf{C}} \triangleq \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$. Unfortunately, this error vector is dependent on $\mathbf{T}$ and $\mathbf{Q}$ which must be removed. According to (14), the matrices $\mathbf{C}$ and $\mathbf{D}$ are orthogonal to $\mathbf{T}$ and $\mathbf{Q}$, respectively. Furthermore,

$$\begin{aligned} \text{rank}(\mathbf{T}) \;+\; \text{rank}(\mathbf{C}) \;&=\; k_1 \qquad \text{and} && (23a) \\ \text{rank}(\mathbf{Q}) \;+\; \text{rank}(\mathbf{D}) \;&=\; k_2 && (23b) \end{aligned}$$

Hence, using a theorem on Projection matrices [14],

$$\begin{aligned} \mathcal{P}_{\mathbf{C}} \;+\; \mathcal{P}_{\mathbf{T}} \;&= \mathbf{I}_{k_1} \qquad \text{and} && (24a) \\ \mathcal{P}_{\mathbf{D}} \;+\; \mathcal{P}_{\mathbf{Q}} \;&= \mathbf{I}_{k_2}. && (24b) \end{aligned}$$

Using these relationships in (22), $\mathcal{P}_{\mathbf{T}}$ and $\mathcal{P}_{\mathbf{Q}}$ can now be replaced and the error vector can be written as,

$$\begin{aligned}
\mathbf{e} &= [(\mathbf{I}_{k_2} - \mathcal{P}_{\mathbf{D}}) \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes (\mathbf{I}_{k_1} - \mathcal{P}_{\mathbf{C}}) + \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}}]\mathbf{x} && (25a)\\
&= [\mathbf{I}_{k_2} \otimes \mathcal{P}_{\mathbf{C}} - \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes \mathbf{I}_{k_1} - \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}}]\mathbf{x} && (25b)\\
&= [\mathbf{I}_{k_2} \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes \mathbf{I}_{k_1} - \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}}]\mathbf{x}. && (25c)
\end{aligned}$$

Note that in this final form of the error, there is no dependence on either $\mathbf{T}$ or $\mathbf{Q}$. Hence, the error criterion for determining the denominator coefficient vectors $\mathbf{c}$ and $\mathbf{d}$ can now be written as,

$$\min_{\mathbf{c},\mathbf{d}} \mathbf{e}\|(\mathbf{c},\mathbf{d})\|^2 = \min_{\mathbf{c},\mathbf{d}} (\mathbf{x}^T[\mathbf{I}_{k_2} \otimes \mathcal{P}_{\mathbf{C}} + \mathcal{P}_{\mathbf{D}} \otimes \mathbf{I}_{k_1} - \mathcal{P}_{\mathbf{D}} \otimes \mathcal{P}_{\mathbf{C}}]\mathbf{x}) \qquad (26)$$

Equations (26) and (9a) represent the desired decoupled criteria for determining the denominator and numerator coefficients, respectively. Optimization of (26) would produce the optimal $\mathbf{c}$ and $\mathbf{d}$, denoted as, $\mathbf{c}^o$ and $\mathbf{d}^o$, respectively. Letting $\mathbf{e}^o$ denote the minimized error corresponding to the optimum denominator coefficients, the optimum spatial-response vector $\mathbf{h}$ can be found from,

$$\mathbf{h}^o \triangleq \mathbf{x} - \mathbf{e}^o. \qquad (27)$$

This $\mathbf{h}^o$ can then be used in (9a) to obtain the optimal numerator vector, $\mathbf{q}^o$.

Analyzing the criteria in (26) it is apparent that the first two terms are the orthogonal projections of the data $\mathbf{x}$ on to the parameter spaces of each of the two spatial dimensions. The third term is the orthogonal projection common to both dimensions but is subtracted once because the common (or, joint) projections have already been included once in each of the first two terms. It is very interesting to note that this criterion is quite analogous to the standard formula of the *Probability of Union* of two subsets. It may be emphasized here that this form of the error criterion is not only mathematically appropriate it is intuitively appealing as well and this form of the 2-D error criterion was not arrived at in any of the previous generalizations of EFM [4, 5]. With further algebraic manipulations, the error-vector $\mathbf{e}$ can also be shown to be related to *both* the denominator vectors $\mathbf{c}$ and $\mathbf{d}$ in a *quasi*-linear manner as,

$$\mathbf{e}(\mathbf{c},\mathbf{d}) \triangleq \left((\mathbf{I}_{k_2} - \mathbf{P}_{\mathbf{D}}) \otimes \mathbf{W}(\mathbf{c}))\mathbf{X}^1 \quad (\mathbf{W}(\mathbf{d}) \otimes \mathbf{I}_{k_1})\mathbf{X}^2\right) \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}, \qquad (28a)$$

where,

$$\begin{aligned}
\mathbf{W}(\mathbf{c}) &\triangleq \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1} \qquad \text{and} && (28b)\\
\mathbf{W}(\mathbf{d}) &\triangleq \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}. && (28c)
\end{aligned}$$

$\mathbf{X}^1$ and $\mathbf{X}^2$ are constructed from the elements in $\mathbf{X}$ as [9, 17],

$$\mathbf{X}^1 \triangleq (\mathbf{X}_1^T \;\; \mathbf{X}_2^T \;\; \cdots \;\; \mathbf{X}_{k_2}^T)^T, \qquad (29)$$

where, the $(l,k)^{th}$ term of $\mathbf{X}_i$ is formed as,

$$\mathbf{X}_i(l,k) \triangleq \mathbf{x}_i(l - k + m_1 + 1), \text{ for } i = 1, 2, \ldots, k_2, \text{ and} \qquad (30a)$$

$$\mathbf{X}^2 \triangleq \begin{pmatrix} \mathbf{x}_{m_2+1} & \cdots & \mathbf{x}_1 \\ \mathbf{x}_{m_2+2} & \cdots & \mathbf{x}_2 \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{k_2} & \cdots & \mathbf{x}_{k_2-m_2} \end{pmatrix}. \qquad (30b)$$

Equation (28a) is one the key results derived in this Section. It clearly shows that both the unknown coefficient vectors $\mathbf{c}$ and $\mathbf{d}$ appear *simultaneously* in a 'quasi-linear' relationship *w.r.t.* the true error vector. This quasi-linear relationship can be exploited for simultaneous optimization of the criterion in (26) *w.r.t.* $\mathbf{c}$ and $\mathbf{d}$. The algorithm is similar in flavor to the ones in [3-5, 8, 13], except that both the denominators are optimized and estimated simultaneously. Specifically, the algorithm iteratively minimizes the ℓ_2-norm of the error vector formed in (26) in two phases. In Phase-1, the $\mathbf{W}$ matrices are treated as constants and are formed using the estimates of $\mathbf{c}$ or $\mathbf{d}$ obtained at the previous iteration. In Phase-2, the estimates are improved upon by setting the gradient of the complete error-norm to zero. The iterations are initialized by setting $\mathbf{c} = [1\ 0\ \ldots\ 0]^T$ and $\mathbf{d} = [1\ 0\ \ldots\ 0]^T$. The iterations are continued until the changes in the estimates in successive iterations become very small. It may be noted here that extensive simulation experience in 1-D [7, 8, 13, 14] as well as for 2-D cases [13, 15] indicate that Phase-1 itself produces very good estimates of the filter coefficients and in most cases, there may not be any need for invoking Phase-2 at all. It may be noted here that in [4, 5], the complete error $\mathbf{e}(\mathbf{c}, \mathbf{d})$ in (26) was not optimized.

Symmetric Spatial Response - A Special Case : Many 2-D filters used in image processing are symmetrically shaped in the spatial domain. Some notable examples are, Gaussian and Circular filters. In designing such spatially symmetric 2-D filters, the methods in [4, 5] sometimes produced slightly different sets of denominator polynomials. Hence, the estimated spatial response may not possess the desired symmetry. This problem may be attributed to separate estimation of the individual denominators. In the proposed approach, both the denominators are optimized simultaneously by minimizing the entire error in (28a). If necessary, the desired symmetry may be imposed by setting, $\mathbf{c} = \mathbf{d}$ in (28a) at the outset. For this special but very important special case, (28a) would have the following form :

$$\mathbf{e}(\mathbf{c}) \underline{\underline{\triangle}} [((\mathbf{I}_{k_2} - \mathbf{P}_\mathbf{C}) \otimes \mathbf{W}_{k_1}(\mathbf{c}))\mathbf{X}^1 + (\mathbf{W}_{k_2}(\mathbf{c}) \otimes \mathbf{I}_{k_1})\mathbf{X}^2]\mathbf{c}, \tag{31}$$

where, the subscripts of $\mathbf{W}$ denote leading dimensions which may be unequal. Minimization of the norm of the error in (31) will result in a *single* set of optimal coefficients meant for both dimensions. This is one of the major advantages of the proposed approach over the ones in [4, 5] where separate optimization in each domain does not necessarily guarantee identical denominator coefficients in both domains.

V. Simulation Results

In order to demonstrate the effectiveness of the proposed algorithm, the results of the design of a Gaussian Filter are given here. The spatial response of a quarter plane Gaussian filter defined over the first quadrant is given by :

$$H(i,j) = 0.256322\ e^{[-0.103203\{(i-4)^2+(j-4)^2\}]},$$

where, $(i,j) \in S_f$ and the support S_f is given by $S_f = \{(i,j) \mid 0 \le i \le 14;\ 0 \le j \le 14\}$. The true or the desired spatial response is shown in Fig. 1. Note that the spatial response is symmetric around its center point. Fig. 3 through 5 show the estimated responses for filter orders $(m_1 = m_2)$, 4, 5 and 6, respectively. The results were obtained by minimizing the norm of the error vector in (31) with different orders. The algorithm converged in 5-7 iterations. The plots clearly show that the estimated spatial impulse responses match the desired one quite closely and, as can be expected, the match improves as the filter order increases. With sixth-order the difference between the true and the estimated response is almost negligible. The closeness between the true and the estimated response was also measured in terms of the ratio of the power of the true response to that of the errors in each case. The ratios were found to be about 41.2dB, 61.4dB and 86.5dB for filter orders 4, 5 and 6, respectively. Simulations with other forms of 2-D filters also showed similar performance.

VI. Conclusion and Future Work :

An optimal 2-D IIR filter design method has been presented. The algorithm is a 2-D extension of an existing optimal 1-D approach. The 2-D model-fitting criterion has been decoupled into a linear and a nonlinear sub-problems. The non-linear part has been shown to possess a quasi-linear relationship with the unknown denominator coefficients. The algorithm simultaneously optimizes the coefficients in both dimensions. Regarding future work, it may be noted that similar to EFM [3], the proposed algorithm is also applicable for strictly-proper designs only, albeit in 2-D. Recently, an optimal 1-D algorithm (OM) which is applicable for any general system with arbitrary number of poles and zeros, has been presented in [14]. Unlike EFM, the general 1-D algorithm in [14] formulates the criterion entirely differently. It shows explicitly that the true error is linearly related to the numerator coefficients whereas the denominator is nonlinearly related. The possibility of extending this work for designing 2-D filters with any arbitrary numbers of denominator and numerator orders is presently under investigation.

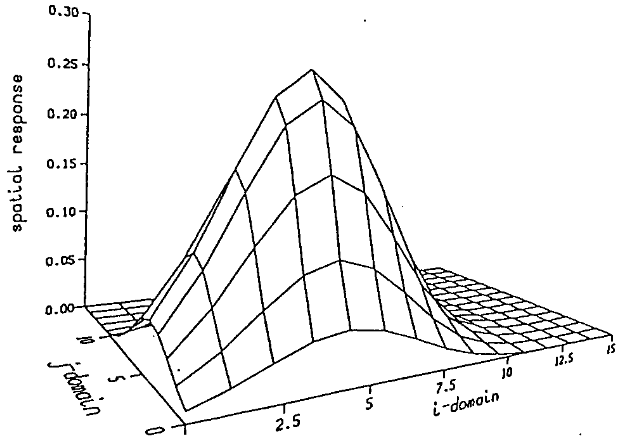


Fig. 1 : The True Spatial-Domain response for a 15 × 15 Gaussian Filter.

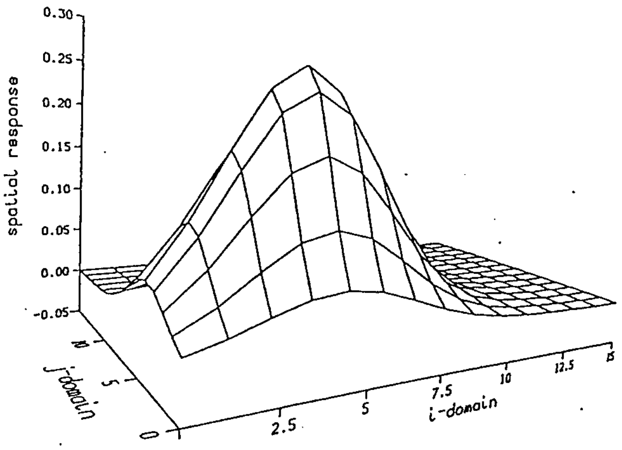


Fig. 2 : The Estimated Spatial-Domain response with $m_1 = n_1 = 4$

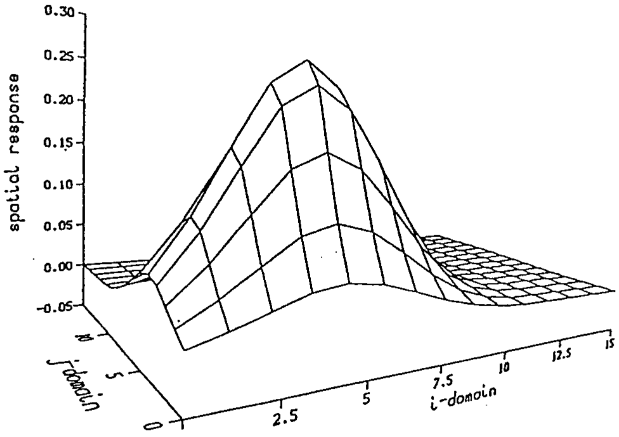


Fig. 3 : The Estimated Spatial-Domain response with $m_1 = n_1 = 5$

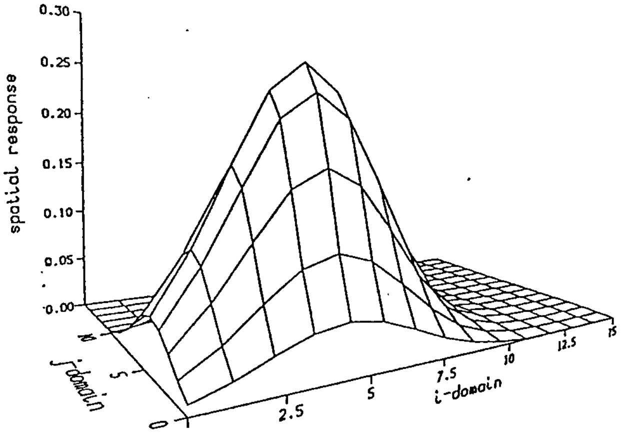
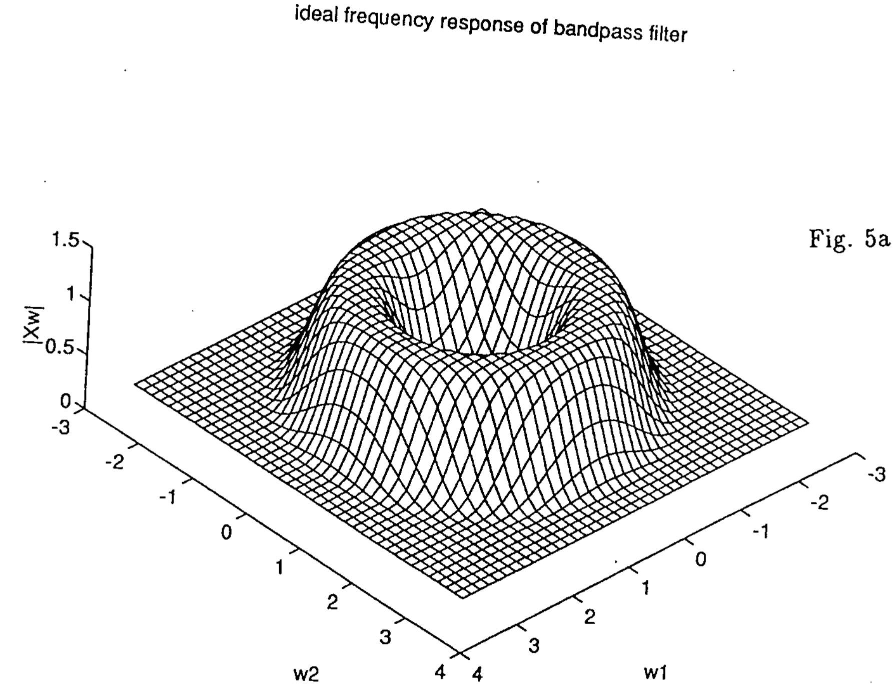
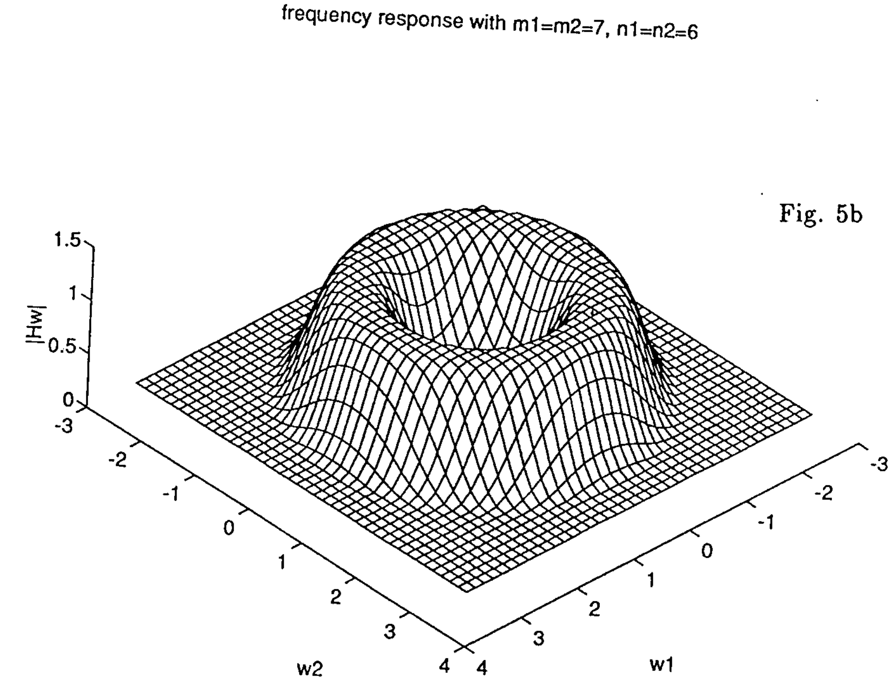


Fig. 4 : The Estimated Spatial-Domain response with $m_1 = n_1 = 6$



ideal frequency response of bandpass filter

Fig. 5a



frequency response with m1=m2=7, n1=n2=6

Fig. 5b

References

[1] J. A. Cadzow, "Recursive Digital Filter Synthesis via Gradient Based Algorithms", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-24, pp. 349-355, June, 1976.

[2] D.E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1984.

[3] A.G. Evans and R. Fischl, "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters", *IEEE Transactions on Audio and Electro-Acoustics*, Vol. AU-21, pp. 61-65, 1973.

[4] T. Hinamoto and S. Maekawa, "Spatial-Domain Design of a Class of Two-Dimensional Recursive Digital Filters", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-32, No.1, Feb., 1984.

[5] T. Hinamoto, "Design of 2-D Separable-Denominator Recursive Digital Filters", *IEEE Transactions on Circuits and Systems*, Vol. CAS-31, pp. 925-933, Nov. 1984.

[6] T. Hinamoto and S. Maekawa, "Separable-Denominator 2-D Rational Approximation via 1-D Based Algorithm", *IEEE Transactions on Circuits and Systems*, Vol. CAS-32, pp. 989-999, Nov. 1985.

[7] C. T. Mullis and R. A. Roberts, "The use of Second-order Information in the Approximation of Discrete-Time Linear Systems," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-24, pp. 226-238, June, 1976.

[8] R. Kumaresan, L. L. Scharf and A. K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol.ASSP-34, pp. 637-640, June, 1986.

[9] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications.* New York; Wiley, 1971.

[10] L. L. Scharf, *Statistical Signal Processing - Detection, Estimation and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[11] J.L.Shanks, "Recursion Filters for Digital Processing", *Geophysics*, Vol. 32, pp. 33-51, 1967.

[12] J.L. Shanks, S. Treitel and J.H. Justice, "Stability and Synthesis of Two-Dimensional Recursive Filters" *IEEE Trans. Audio Electroacoust.*, Vol. AU-20, pp. 115-128, 1972.

[13] A.K. Shaw, *Structured Matrix Problems in Signal Processing*, Ph.D. Dissertation, Univ. of Rhode Island, RI, 1987.

[14] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," to appear, *IEEE Transactions on Signal Processing*, Jan., 1994.

[15] A.K. Shaw and R. Kumaresan, "Some Structured Matrix Approximation Problems", *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, New York, NY, pp. 2324-2327, April, 1988.

[16] G. A. Shaw and R. M. Mersereau, "Design, Stability and Performance of Two-Dimensional Recursive Digital Filters", *Tech. Report E21-B05-1*, Georgia Inst. of Technology School of Electrical Engg., 1979.

[17] K. Steiglitz and L.E. McBride, "A Technique for Identification of Linear Systems", *IEEE Transactions on Automatic Control*, Vol. AC-10, pp. 461-464, 1965.

Section - 3.5 : Optimal Frequency Domain Design of Denominator Separable Two-Dimensional Digital IIR Filters

SUMMARY

Classical design techniques using Butterworth, Chebyshev or Elliptic polynomial are only limited particular types of design specifications, such as Bandpass, lowpass etc. A least-squares technique is presented for designing quarter-plane separable-denominator 2-D IIR filters to best approximate prescribed frequency domain (FD) specification of any arbitrary shape. Structured Matrix Approximation approach is utilized to show that the FD error vector is linearly related to the 2-D numerator coefficients whereas the relationship with the 2-D denominators is quasi-linear. Furthermore, the numerator and denominator estimation problems are theoretically decoupled. The quasi-linear relationship with the denominator is used to formulate an algorithm for iterative estimation of the denominator. The numerator is found in one step using the estimated denominator. Computer simulations show the effectiveness of the proposed method and its superior performance compared to several existing methods.

I. Introduction :

Design of 2-D digital IIR filters from arbitrary frequency domain specifications is a highly nonlinear optimization problem [1-10]. Existing designs make use of variations of general nonlinear optimization techniques, such as Newton-Raphson or Fletcher-Powell or linear programming to meet the prescribed design specifications [3, 4, 6-8]. But these general methods are computationally intensive, highly sensitive to the choice of initial estimates and may take large number of iterations. Also, none of these methods make use of the underlying matrix-structure inherent in the 2-D filter design problem. For Spatial Domain designs, it has been shown by several researchers (including the second author) that appropriate utilization of the underlying matrix structures leads to insightful theoretical framework and efficient computational algorithms [1, 2, 5, 9, 10]. In practice though, the filter specifications are usually prescribed in the frequency-domain and hence, direct design in the frequency domain would certainly be more desirable. To the best knowledge of the authors no structured matrix framework has been developed for frequency-domain 2D IIR filter design. The primary goal of this work is to fill this gap by demonstrating that an equivalent structured matrix framework does exist in the frequency-domain also and furthermore, it can be utilized equally effectively for designing 2D IIR filters. Though the proposed framework can be adapted for general cases, we present the design of denominator-separable filters here because the inherent symmetry in many commonly used 2D filters conform to the separable-denominator structure and the stability of these filters can be easily verified.

This work shows that the optimal 2D rational model identification problem belongs to a special class of *mixed*-nonlinear optimization problem where the linear and nonlinear parameters appear separately. Furthermore, the mixed nonlinear criterion can be decoupled into a *purely* linear problem for estimating the numerator and a separate nonlinear problem of reduced dimensionality, for estimating the separable denominators. The matrix structure of the nonlinear denominator criterion naturally leads to an iterative algorithm whereas the numerator is estimated with a single step of Least-Squares estimation. In simulations, the proposed approach provides superior match than various existing general approaches.

II. Problem Definition :

The transfer function of a 2-D separable-denominator LSI system is given by

$$H(z_1, z_2) = \frac{A(z_1, z_2)}{B(z_1)\, C(z_2)} = \frac{\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} a(i,j) z_1^{-i} z_2^{-j}}{\sum_{i=0}^{m_1} b(i) z_1^{-i} \, \sum_{j=0}^{m_2} c(j) z_2^{-j}} = \frac{\mathbf{z}_1^T \mathbf{A} \mathbf{z}_2}{\mathbf{z}_1^T \mathbf{b} \mathbf{c}^T \mathbf{z}_2} \tag{1}$$

where, $\mathbf{b} \triangleq [b(0)\ b(1) \cdots b(m_1)]^T$, $\mathbf{c} \triangleq [c(0)\ c(1) \cdots c(m_2)]^T$,

$$\mathbf{A} \triangleq \begin{bmatrix} a(0,0) & a(0,1) & \cdots & a(0,m_2) \\ a(1,0) & a(1,1) & \cdots & a(1,m_2) \\ \vdots & \vdots & \ddots & \vdots \\ a(m_1,0) & a(m_1,1) & \cdots & a(m_1,m_2) \end{bmatrix}. \tag{2}$$

and $\mathbf{z}_1$ and $\mathbf{z}_2$ are vectors of the form $\mathbf{z}_i \triangleq [1\ z_i^{-1}\ z_i^{-2} \cdots]^T$ with appropriate sizes. Let the $k_1 \times k_2$ *desired* frequency response be

$$\mathbf{X}_d \triangleq \begin{bmatrix} x(\omega_{11},\omega_{21}) & x(\omega_{11},\omega_{22}) & \cdots & x(\omega_{11},\omega_{2k_2}) \\ \vdots & \vdots & \ddots & \vdots \\ x(\omega_{1k_1},\omega_{21}) & x(\omega_{1k_1},\omega_{22}) & \cdots & x(\omega_{1k_1},\omega_{2k_2}) \end{bmatrix} \tag{3}$$

and the frequency response of the separable-denominator filter at the same frequency points be $\mathbf{X}$. Let $\mathbf{x}_d \triangleq vec(\mathbf{X}_d)$ and $\mathbf{x} \triangleq vec(\mathbf{X})$. The problem is to estimate the coefficients in $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{A}$ by optimizing the following 2-D least-squares error criterion,

$$\min_{\mathbf{b},\mathbf{c},\mathbf{A}} \|\mathbf{e}\|^2 \triangleq \|\mathbf{x}_d - \mathbf{x}\|^2 \qquad \text{with } b(0) = 1, c(0) = 1. \tag{4}$$

III. Decoupling the error-criterion :

Let $H_b(z_1)$ and $H_c(z_2)$ be the inverse filters of $B(z_1)$ and $C(z_2)$ respectively *i.e.,* $B(z_1)H_b(z_1) = 1$ and $C(z_2)H_c(z_2) = 1$. The system function can therefore be written as

$$\begin{aligned} H(z_1,z_2) &= \frac{A(z_1,z_2)}{B(z_1)\,C(z_2)} \\ &= \frac{\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} a(i,j) z_1^{-i} z_2^{-j}}{\sum_{i=0}^{m_1} b(i) z_1^{-i}\ \sum_{j=0}^{m_2} c(j) z_2^{-j}} = H_b(z_1)\,A(z_1,z_2)\,H_c(z_2) \end{aligned} \tag{5}$$

Assuming $k_1 \times k_2$ significant samples for the spatial response, the above relation can be expressed in matrix notation as

$$\mathbf{H} = \mathbf{H}_L^b \mathbf{A} \mathbf{H}_L^{cT} \tag{6}$$

where,

$$\mathbf{H} \triangleq \begin{bmatrix} h(0,0) & h(0,1) & \cdots & h(0,k_2-1) \\ h(1,0) & h(1,1) & \cdots & h(1,k_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(k_1-1,0) & h(k_1-1,1) & \cdots & h(k_1-1,k_2-1) \end{bmatrix}, \tag{7}$$

$$\mathbf{H}_L^b(i,j) \triangleq \begin{bmatrix} h_b(0) & 0 & \cdots & 0 \\ h_b(1) & h_b(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_b(n_1) & h_b(n_1-1) & \cdots & h_b(0) \\ \vdots & \vdots & \ddots & \vdots \\ h_b(k_1-1) & h_b(k_1-2) & \cdots & h_b(k_1-n_1-1) \end{bmatrix}, \tag{8}$$

$$\mathbf{H}_L^c \triangleq \begin{bmatrix} h_c(0) & 0 & \cdots & 0 \\ h_c(1) & h_c(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_c(n_2) & h_c(n_2-1) & \cdots & h_b(0) \\ \vdots & \vdots & \ddots & \vdots \\ h_c(k_2-1) & h_c(k_2-2) & \cdots & h_c(k_2-n_2-1)) \end{bmatrix}, \tag{9}$$

The frequency response of the 2-D filter can be written in a matrix-decomposed form as,

$$\mathbf{X} = \mathbf{W}_b \mathbf{H} \mathbf{W}_c^T \tag{10}$$

where,

$$\mathbf{W}_b \triangleq \begin{bmatrix} 1 & e^{-j\omega_{11}} & e^{-j2\omega_{11}} & \cdots & e^{-j(k_1-1)\omega_{11}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\omega_{1k_1}} & e^{-j2\omega_{1k_1}} & \cdots & e^{-j(k_1-1)\omega_{1k_1}} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_c \triangleq \begin{bmatrix} 1 & e^{-j\omega_{21}} & e^{-j2\omega_{21}} & \cdots & e^{-j(k_2-1)\omega_{21}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\omega_{2k_2}} & e^{-j2\omega_{2k_2}} & \cdots & e^{-j(k_2-1)\omega_{2k_2}} \end{bmatrix}. \tag{11}$$

Applying the *vec* operator on both sides of (10), we get

$$vec(\mathbf{X}) \triangleq \mathbf{x} = vec(\mathbf{W}_b \mathbf{H} \mathbf{W}_c^T) = vec(\mathbf{W}_b \mathbf{H}_L^b \mathbf{A} \mathbf{H}_L^{cT} \mathbf{W}_c^T) = (\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b) vec(\mathbf{A}) = (\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b)\mathbf{a}. \tag{12}$$

Hence, the error between the desired and the filter frequency response, as defined in (3), can be written as,

$$\mathbf{e} = \mathbf{x}_d - \mathbf{x} = \mathbf{x}_d - (\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b)\mathbf{a}. \tag{13}$$

This expression shows explicitly that the frequency domain error is linearly related to the numerator coefficient vector **a** and nonlinearly related to the denominators in a rather complicated manner. Interestingly, if the denominator coefficients are known, the least-squares estimate of the numerator coefficients can be obtained by minimizing (4),

$$\mathbf{a} = (\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b)^{\#} \mathbf{x}_d \tag{14}$$

where # denotes the pseudo-inverse of the matrix. Substituting this in (10), we get the *decoupled* denominator criterion,

$$||\mathbf{e}(\mathbf{b},\mathbf{c})||^2 = ||\mathbf{x}_d - (\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b)(\mathbf{W}_c \mathbf{H}_L^c \otimes \mathbf{W}_b \mathbf{H}_L^b)^{\#} \mathbf{x}_d||^2 = ||(\mathbf{I}_{k_1 k_2} - (\mathbf{P}_{\mathbf{W}_c \mathbf{H}_L^c} \otimes \mathbf{P}_{\mathbf{W}_b \mathbf{H}_L^b}))\mathbf{x}_d||^2 \tag{15}$$

where, $\mathbf{P}_\mathbf{Y} \triangleq \mathbf{Y}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H$ denotes the projection matrix of a matrix $\mathbf{Y}$ with H being the conjugate-transpose operator. Extending Theorem 2.1 in [12], it can be shown that if the denominator is estimated by minimizing criterion in (15) and that estimate is used in (14), the estimates retain the global optima of the original criterion in (4).

IV. Reparametrization of the error-criterion :

In this section the *decoupled* criterion in (12) will be directly related to the denominator coefficients. The inverse filter relation $B(z_1)H_b(z_1) = 1$ can be expressed in matrix notation as

$$\mathbf{B}_L \mathbf{H}_b = \mathbf{I}_{k_1} \tag{16}$$

$$\text{where,}\quad \mathbf{B}_L \triangleq \left[\begin{array}{ccccccccc} b(0) & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \\ \hdashline b(n_1+1) & \cdots & \cdots & b(0) & 0 & \cdots & \cdots & \cdots & \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \\ b(m_1) & \cdots & \cdots & \cdots & b(0) & 0 & \cdots & 0 & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \cdots & \cdots & b(m_1) & \cdots & \cdots & \cdots & b(0) & \end{array}\right] \triangleq \left[\begin{array}{c} \mathbf{B}_u \\ --- \\ \mathbf{B}^T \end{array}\right] \tag{17}$$

$$\text{and}\quad \mathbf{H}_b \triangleq \left[\begin{array}{cccc|ccc} h_b(0) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ h_b(n_1) & h_b(n_1-1) & \cdots & h_b(0) & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h_b(k_1-1) & h_b(k_1-2) & \cdots & h_b(k_1-n_1-1) & h_b(k_1-n_1-2) & \cdots & h_b(0) \end{array}\right] \triangleq [\mathbf{H}_L^b \;\mid\; \mathbf{H}_R^b] \tag{18}$$

Let, $\mathbf{W}_{b_I}\mathbf{W}_b = \mathbf{I}_{k_1}$. This inverse exists because the frequencies ω_{1k}'s are distinct. Using it in (16) with (17) and (18),

$$\mathbf{B}\mathbf{W}_{b_I}\mathbf{W}_b\mathbf{H}_b = \mathbf{I}_{k_1} = \left[\begin{array}{c} \mathbf{B}_u \\ --- \\ \mathbf{B}^T \end{array}\right] \mathbf{W}_{b_I}\mathbf{W}\,[\mathbf{H}_L^b \;\mid\; \mathbf{H}_R^b] = \left[\begin{array}{c|c} \mathbf{B}_u\mathbf{W}_{b_I}\mathbf{W}\mathbf{H}_L^b & \mathbf{B}_u\mathbf{W}_{b_I}\mathbf{W}\mathbf{H}_R^b \\ ---------- & ---------- \\ \mathbf{B}^T\mathbf{W}_{b_I}\mathbf{W}\mathbf{H}_L^b & \mathbf{B}^T\mathbf{W}_{b_I}\mathbf{W}\mathbf{H}_R^b \end{array}\right] \tag{19}$$

The bottom-left corner element of the matrix at right suggests that $\mathbf{W}_{b_I}^H\mathbf{B}$ and $\mathbf{W}\mathbf{H}_L^b$ are orthogonal, *i.e.*, $(\mathbf{W}_{b_I}^H\mathbf{B})^H(\mathbf{W}\mathbf{H}_L^b) = \mathbf{0}$. Also, since $rank(\mathbf{W}_{b_I}^H\mathbf{B}) + rank(\mathbf{W}\mathbf{H}_L^b) = (k_1 - n_1 - 1) + (n_1 + 1) = k_1$, using a theorem on projection matrices,

$$\mathbf{P}_{\mathbf{W}_{b_I}^H\mathbf{B}} + \mathbf{P}_{\mathbf{W}\mathbf{H}_L^b} = \mathbf{I}_{k_1}. \tag{20}$$

Similarly, from the inverse filter relation $C(z_2)H_c(z_2) = 1$, we can get

$$\mathbf{P}_{\mathbf{W}_{c_I}^H\mathbf{C}} + \mathbf{P}_{\mathbf{W}\mathbf{H}_L^c} = \mathbf{I}_{k_2}. \tag{21}$$

Substituting the above relations in (12) and using Kronecker product representation ($\otimes$), the error can be written as

$$\mathbf{e}(\mathbf{b},\mathbf{c}) \triangleq [(\mathbf{I}_{k_2} - \mathbf{P}_{\mathbf{W}_{c_I}^H\mathbf{C}}) \otimes \mathbf{P}_{\mathbf{W}_{b_I}^H\mathbf{B}} + \mathbf{P}_{\mathbf{W}_{b_I}^H\mathbf{B}} \otimes \mathbf{I}_{k_1}]\,\mathbf{x}_d = [((\mathbf{I}_{k_2} - \mathbf{P}_{\mathbf{W}_{c_I}^H\mathbf{C}}) \otimes \mathbf{V}_b)\mathbf{X}^1 \;\; (\mathbf{V}_c \otimes \mathbf{I}_{k_1})\,\mathbf{X}^2] \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix}. \text{ where,} \tag{22}$$

$\mathbf{X}^1$ and $\mathbf{X}^2$ are formed with prescribed data, $\mathbf{V}_b \triangleq (\mathbf{W}_{b_I}^H\mathbf{B})(\,(\mathbf{W}_{b_I}^H\mathbf{B})^H\,(\mathbf{W}_{b_I}^H\mathbf{B})\,)^{-1}$ and $\mathbf{V}_c \triangleq (\mathbf{W}_{c_I}^H\mathbf{C})(\,(\mathbf{W}_{c_I}^H\mathbf{C})^H\,(\mathbf{W}_{c_I}^H\mathbf{C})\,)^{-1}$.

V. Simulation Results :

Several designs were implemented using the proposed algorithm and the performances were compared with existing approaches. Fig. 1-3 show the results of the comparison. Fig. 1a, 2a and 3a show results using the methods proposed in [6], [7] and [8], respectively. For the same or less numerator/denominator orders, Fig. 1b, 2b and 3b show the corresponding results using the proposed method. The relative rms errors [7] for the results in Fig. 1a, 2a and 3a are 0.67, 0.28 and 0.77, respectively. The errors for the results in Fig. 1b, 2b and 3b are 0.21, 0.26 and 0.68 respectively. Clearly, the proposed approach found better match with lesser number of coefficients, in all cases. The number of iterations for the proposed approach were less than 10 in all cases, whereas the general optimization approaches sometimes took close to hundred or more iterations.

Other Methods

Proposed Method

Fig. 1a : Method in [6]

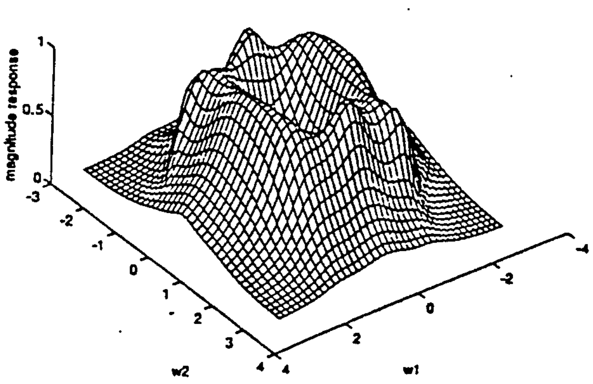


Fig. 1b

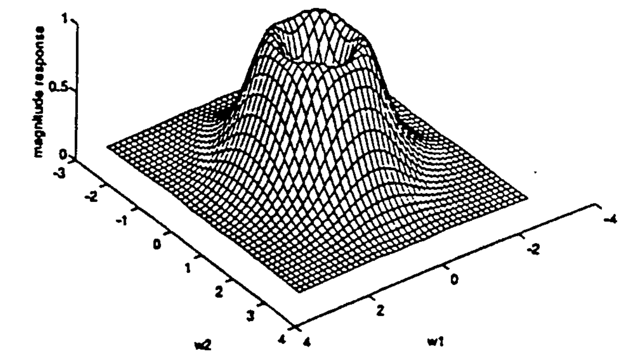


Fig. 2a : Method in [7]

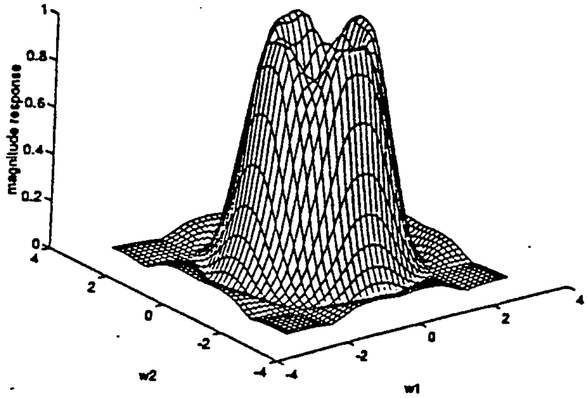


Fig. 2b

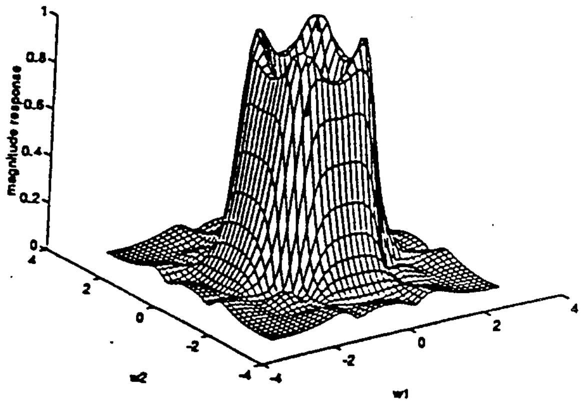


Fig. 3a : Method in [8]

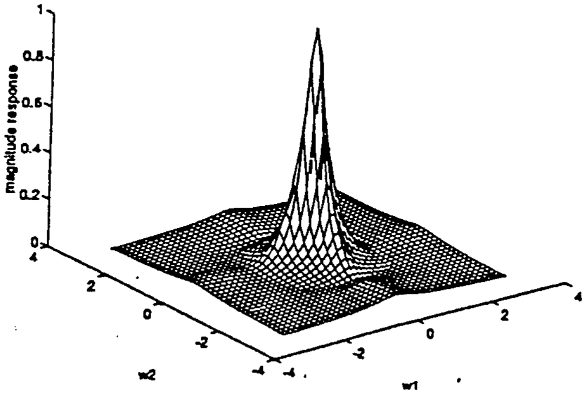
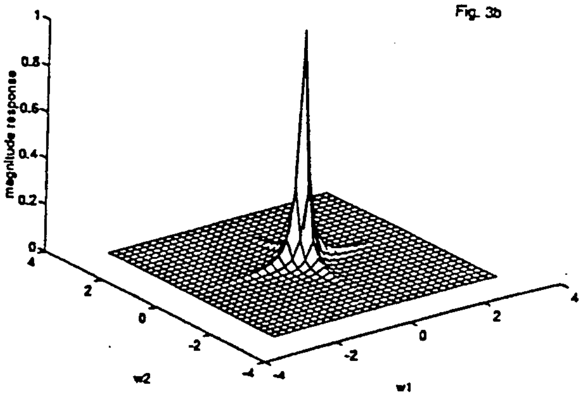


Fig. 3b



References

[1] J.L. Shanks, S. Treitel and J.H. Justice, "Stability and Synthesis of Two-Dimensional Recursive Filters" *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-20, pp. 115-128, 1972.

[2] T. Hinamoto and S. Maekawa, "Spatial-Domain Design of a Class of Two-Dimensional Recursive Digital Filters", *IEEE Trans. on ASSP*, Vol.- ASSP-32, no.1, Feb.,1984.

[3] T. Hinamoto and F.W. Fairman, "Separable Denominator State Space Realization of Two Dimensional Filters using a Canonical Form", *IEEE Trans. Acoust. Speech, Sig. Proc.*, vol. ASSP-29, no. 4, pp. 846-853, 1981.

[4] T. Hinamoto and S. Maekawa, "Separable-Denominator 2-D Rational Approximation via 1-D Based Algorithm ", *IEEE Trans. Cir.Syst.*, Vol. CAS-32, pp. 989-999, Nov. 1985.

[5] T. Hinamoto, "Design of 2-D Separable-Denominator Recursive Digital Filters", *IEEE Trans. Cir.Syst.*, Vol. CAS-31, pp. 925-933, Nov. 1984.

[6] A. T. Chottera and G. A. Jullien, "Design of Two-Dimensional Recursive Digital Filters Using Linear Programming", *IEEE Trans. Cir. Syst.*, CAS-29, pp. 817-826, Dec. 1982.

[7] T. Hinamoto and S. Maekawa, "Design of Two-Dimensional Recursive Digital Filters using Mirror Image Polynomials", *IEEE Trans. Cir. Syst.*, CAS-33, pp. 750-758, Aug. 1986.

[8] S. A. H. Aly and M. M. Fahmy, "Design of Two-Dimensional Recursive Digital Filters with Specified Group Delay Characteristics," *IEEE Trans. Cir. Syst.*, CAS-25, pp. 908-915, Nov. 1978.

[9] A. K. Shaw and P. Misra, "Optimal 2-D IIR Filters : Strictly Proper Case ," *ICASSP-92*, San Francisco, CA, IV-333-336, April, 1992.

[10] A. K. Shaw, "Design of Denominator Separable 2-D IIR Filters," *Signal Processing,* to be published, Vol. 2, 1994.

[11] A. K. Shaw, "Optimal Design of Digital Digital IIR Filters by Model Fitting Frequency Response Data," *IEEE Trans. Cir. Sys.*, accepted, 1994.

[12] G. H. Golub and V. Pereyra, "The Differentiation of Pseudoinverses and Nonlinear Problems Whose Variables Separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413-432, Apr., 1973.

Section - 3.6 : OPTIMAL SPATIAL-DOMAIN DESIGN OF 2-D IIR FILTERS

SUMMARY

In this Section we present a structured matrix approximation framework to develop the most general form for *optimal* least-squares (LS) design of 2-D recursive filters from prescribed spatial domain data. Unlike the work in Section 3.4, no separability is assumed for the 2D denominator. Utilizing matrix structures inherent in this problem it is shown that the exact ℓ_2 error has a purely *linear* relationship with the 2-D numerator parameters whereas the 2-D denominator coefficients are nonlinearly related to the error. But more interestingly, the denominator and numerator estimation problems are theoretically decoupled into separate problems without affecting any optimality properties. In the decoupled form, the numerator estimation problem is shown to be purely linear. For estimating the denominator also, it is shown that the decoupled ℓ_2 error vector possesses a *quasi-linear* relationship with the denominator coefficients. Decoupled estimation leads to reduced computational complexity because there is no need for iterating on the numerators. Simulation results indicate that for several common filer design problems, the proposed general version performs better than the separable design developed earlier in Section 3.4.

Introduction

Many 2-D filter synthesis algorithms have been developed by extending existing algorithms for 1-D filter design. Specifically, Shanks *et al* [1] extended the work of Shanks [5]; Cadzow [2] and Shaw and Mersereau [3] utilized many of the general non-linear optimization methods; and Shaw and Mersereau [3] also extended the work of Steiglitz and McBride [6]. But these methods are *suboptimal* in the sense that they do not optimize the exact fitting error criterion. In contrast to these approaches, the iterative method (EFM) proposed by Evans and Fischl [7] is *optimal* in the 1-D case because it does optimize the true error criterion. There have been some previous attempts in generalizing EFM to 2-D also [10-13] but, as shown in this paper, the complete error criterion for the most general case has not yet been developed or optimized. Even the suboptimal error criterion had not not optimized *w.r.t.* the filter coefficients in two dimensions simultaneously. The Evans-Fischl method has been found to be highly accurate for 1-D filter design. In a recent work, we have extended 1-D EFM to designing 2-D IIR filters with separable denominators [8], where, unlike several existing 2-D methods [1,3,10-13], the *exact* fitting error was minimized *w.r.t.* the filter coefficients in both dimensions *simultaneously*. Simultaneous optimization was shown to be effective for some commonly occurring design problems with symmetric spatial response.

In this paper we present a structured matrix approximation framework to develop the most general 2-D version of EFM for *optimal* least-squares (LS) design of 2-D recursive filters from prescribed spatial domain data. Utilizing matrix structures inherent in this problem it is shown that the exact ℓ_2 error has a purely *linear* relationship with the 2-D numerator parameters whereas the 2-D denominator coefficients are nonlinearly related to the error. But more interestingly, these two sets of parameters appear separately in the 2-D LS criterion. Hence, using a theorem on separability from Numerical Analysis literature, it is shown that the numerator and denominator estimation problems can be mathematically *decoupled* without affecting any optimality properties. In the decoupled form, the numerator estimation problem is shown to be purely linear. For estimating the denominator also, it is shown that the decoupled ℓ_2 error vector possesses a *quasi-linear* relationship with the denominator coefficients. Decoupled estimation leads to reduced computational complexity because there is no need for iterating on the numerators. Simulation results indicate that for several common filer design problems, the proposed general version performs better than the separable design developed earlier by the Principal Investigator.

Formulation of the Problem

Let the *prescribed* first-quadrant spatial impulse response of size $k_1 \times k_2$ be given by

$$\mathbf{X} = \begin{bmatrix} x(0,0) & x(0,1) & \cdots & x(0,k_2-1) \\ x(1,0) & x(1,1) & \cdots & x(1,k_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ x(k_1-1,0) & x(k_1-1,1) & \cdots & x(k_1-1,k_2-1) \end{bmatrix}. \tag{1}$$

Let

$$H(z_1,z_2) = \frac{A(z_1,z_2)}{B(z_1,z_2)} = \frac{\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} a(i,j) z_1^{-i} z_2^{-j}}{\sum_{i=0}^{m_1}\sum_{j=0}^{m_2} b(i,j) z_1^{-i} z_2^{-j}} \tag{2}$$

be the transfer function of the 2-D filter to be designed to approximate $\mathbf{X}$ and let its $k_1 \times k_2$ spatial response be given by

$$\mathbf{H} = \begin{bmatrix} h(0,0) & h(0,1) & \cdots & h(0,k_2-1) \\ h(1,0) & h(1,1) & \cdots & h(1,k_2-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(k_1-1,0) & h(k_1-1,1) & \cdots & h(k_1-1,k_2-1) \end{bmatrix}. \tag{3}$$

In order to develop the structured-matrix representation of the 2-D filter design problem, define two matrices containing the numerator and denominator coefficients as

$$\mathbf{A} = \begin{bmatrix} a(0,0) & a(0,1) & \cdots & a(0,n_2) \\ a(1,0) & a(1,1) & \cdots & a(1,n_2) \\ \vdots & \vdots & \ddots & \vdots \\ a(n_1,0) & a(n_1,1) & \cdots & a(n_1,n_2) \end{bmatrix} \in \mathbb{R}^{(n_1+1)\times(n_2+1)}, \quad \text{and} \tag{4}$$

$$\mathbf{B} = \begin{bmatrix} b(0,0) & b(0,1) & \cdots & b(0,m_2) \\ b(1,0) & b(1,1) & \cdots & b(1,m_2) \\ \vdots & \vdots & \ddots & \vdots \\ b(m_1,0) & b(m_1,1) & \cdots & b(m_1,m_2) \end{bmatrix} \in \mathbb{R}^{(m_1+1)\times(m_2+1)}, \tag{5}$$

respectively. In vector form define, $\mathbf{x} = vec(\mathbf{X})$, $\mathbf{h} = vec(\mathbf{H})$, $\mathbf{a} = vec(\mathbf{A})$ and $\mathbf{b} = vec(\mathbf{B})$ where *vec* is the operation of stacking all the columns of a matrix one below the other. The problem addressed in this paper is to estimate $\mathbf{a}$ and $\mathbf{b}$ by minimizing the following ℓ_2-norm of the error between $\mathbf{x}$ and $\mathbf{h}$ i.e.,

$$\min_{\mathbf{a},\mathbf{b}} \| \mathbf{e} \|^2 \ \underline{\triangle}\ \| \mathbf{x} - \mathbf{h} \|^2 \tag{6}$$

Decoupling the Error Criterion

Equation (2) can be rewritten as

$$A(z_1,z_2) = H(z_1,z_2)\, B(z_1,z_2). \tag{7}$$

Note that $k_1 \times k_2$ *significant* samples of the desired spatial response are to be matched. Hence, by equating on both sides of (7) the coefficients of equal powers of z_1^{-1} and z_2^{-1} up to $k_1 - 1$ and $k_2 - 1$, respectively, equation (7) can be expressed using matrix-vector notation as

$$\begin{bmatrix} \mathbf{a} \\ \cdots \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^1 \\ \cdots \\ \mathbf{H}^2 \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{B}^1 \\ \cdots \\ \mathbf{B}^T \end{bmatrix} \mathbf{h}, \tag{8}$$

where $\mathbf{H}^1$, $\mathbf{H}^2$, $\mathbf{B}^1$ and $\mathbf{B}^T$ are defined in the Appendix. If the impulse response $\mathbf{H}$ and the numerator coefficients $\mathbf{b}$ were known, the numerator coefficients $\mathbf{a}$ can be calculated from the top part of (8) as

$$\mathbf{a} = \mathbf{H}^1\,\mathbf{b} = \mathbf{B}^1\,\mathbf{h}. \tag{9}$$

However, in this case the exact $\mathbf{H}$ is unknown. Hence, replacing $\mathbf{H}^2$ in the lower half of (8) by $\mathbf{X}^2$ formed using the corresponding elements of $\mathbf{X}$, would produce the following 'equation error',

$$\mathbf{d}(\mathbf{b}) \overset{\triangle}{=} \mathbf{X}^2\,\mathbf{b} = \mathbf{B}^T\,\mathbf{x}. \tag{10}$$

From (6), $\mathbf{x} = \mathbf{h} + \mathbf{e}$. Using this in (10), we get

$$\begin{aligned} \mathbf{d}(\mathbf{b}) =& \mathbf{B}^T\,(\mathbf{h} + \mathbf{e}) \\ =& \mathbf{B}^T\,\mathbf{e}, \quad \text{using (8)} \end{aligned} \tag{11}$$

Also according to (8), $\mathbf{B}$ is orthogonal to $\mathbf{h}$. Hence, based on the orthogonality principle [9] an inverse relationship can be established using similar steps as in [7, 8],

$$\begin{aligned} \mathbf{e} =& \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x} \\ \overset{\triangle}{=}& \ \mathbf{W}\mathbf{B}^T\mathbf{x} \\ =& \mathbf{W}\mathbf{X}^2\mathbf{b} \\ \overset{\triangle}{=}& \ \mathbf{W}\begin{bmatrix}\mathbf{g} & \vdots & \mathbf{G}\end{bmatrix}\mathbf{b} \\ =& \mathbf{W}\mathbf{g} + \mathbf{W}\mathbf{G}\hat{\mathbf{b}} \end{aligned} \tag{12}$$

where $\mathbf{b} \overset{\triangle}{=} [1 \vdots \hat{\mathbf{b}}^T]^T$, $\mathbf{g}$ is the first column of $\mathbf{X}^2$ and $\mathbf{G}$ contains the remaining columns of $\mathbf{X}^2$. If $\mathbf{W}$ is assumed to be independent of $\mathbf{b}$, by setting the gradient of $||\mathbf{e}||^2$ in (13) to zero, the denominator vector $\mathbf{b}$ can be estimated as,

$$\hat{\mathbf{b}} = -(\mathbf{G}^T\mathbf{W}^T\mathbf{W}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}^T\mathbf{W}\mathbf{g} \tag{13}$$

But since $\mathbf{W}$ does depend on $\mathbf{b}$, the above equation will be used to estimate $\mathbf{b}^{(i)}$ iteratively, with $\mathbf{W}$ formed using $\mathbf{b}^{(i-1)}$ estimated at the previous iteration. A convenient initial estimate of $\mathbf{b}$ can be obtained by minimizing the equation error in (10) as

$$\hat{\mathbf{b}}^{(0)} = \begin{bmatrix} 1 \\ \cdots\cdots \\ -(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T \end{bmatrix} \tag{13}$$

Simulation Results

Computer simulations have been done to design zero-phase lowpass [14], zero-phase bandpass [14], Gaussian and Laplacian filters. It has been found that for same filter orders, the proposed method performs better than the separable-denominator case [8] for the lowpass and the bandpass cases, while for the Gaussian and the Laplacian cases the separable-denominator method [8] appears to perform better. Also, in all the examples, the proposed method performed better than the modified Prony's method given in [14]. Some plots have been included of the obtained designs.

Appendix

Definitions of the matrices used in (8) are as follows. It should be noted that the matrix structures have been given only for, $n_1 + 1 \leq m_1$, $n_2 + 1 \leq m_1$, the modifications for other cases being obvious.

$$\mathbf{H}^1 \triangleq \begin{bmatrix} \mathbf{H}_0^1 & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{H}_1^1 & \mathbf{H}_0^1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{H}_{n_2}^1 & \mathbf{H}_{n_2-1}^1 & \cdots & \mathbf{H}_0^1 & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(n_1+1)(n_2+1)\times(m_1+1)(m_2+1)}, \text{ where,} \tag{A.1}$$

$$\mathbf{H}_i^1 \triangleq \begin{bmatrix} h(0,i) & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ h(1,i) & h(0,i) & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ h(n_1,i) & h(n_1-1,i) & \cdots & h(0,i) & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(n_1+1)\times(m_1+1)}, \text{ and} \tag{A.2}$$

$$\mathbf{H}^2 \triangleq \begin{bmatrix} \mathbf{J}_0^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{J}_1^2 & \mathbf{J}_0^2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{m_2}^{m_2+1} & \mathbf{J}_{m_2-1}^{m_2+1} & \cdots & \mathbf{J}_0^{m_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{k_2-1}^{k_2} & \mathbf{J}_{k_2-2}^{k_2} & \cdots & \mathbf{J}_{k_2-m_2-1}^{k_2} \end{bmatrix} \in \mathbb{R}^{[k_1k_2-(n_1+1)(n_2+1)]\times(m_1+1)(m_2+1)}, \tag{A.3}$$

$$\mathbf{J}_i^j \triangleq \begin{bmatrix} h(n_1+1,i) & \cdots & h(0,i) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h(m_1,i) & \cdots & \cdots & \cdots & \cdots & h(0,i) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h(k_1-1,i) & \cdots & \cdots & \cdots & \cdots & h(k_1-m_1-1,i) \end{bmatrix} \in \mathbb{R}^{(k_1-n_1-1)\times(m_1+1)}, \text{ for } j \leq (n_2+1) \tag{A.4}$$

and

$$\mathbf{J}_i^j \triangleq \begin{bmatrix} h(0,i) & 0 & \cdots & 0 \\ h(1,i) & h(0,i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(m_1,i) & h(m_1-2,i) & \cdots & h(0,i) \\ \vdots & \vdots & \ddots & \vdots \\ h(k_1-1,i) & h(k_1-2,i) & \cdots & h(k_1-m_1-1,i) \end{bmatrix} \in \mathbb{R}^{k_1\times(m_1+1)}, \text{ for } j > (n_2+1) \tag{A.5}$$

$$\mathbf{B}^1 \triangleq \begin{bmatrix} \mathbf{B}_0^1 & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{B}_1^1 & \mathbf{B}_0^1 & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{B}_{n_2}^1 & \cdots & \cdots & \mathbf{B}_0^1 & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(n_1+1)(n_2+1)\times k_1k_2}, \text{ where,} \tag{A.6}$$

$$\mathbf{B}_i^1 \triangleq \begin{bmatrix} b(0,i) & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ b(1,i) & b(0,i) & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ b(n_1,i) & \cdots & \cdots & b(0,i) & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(n_1+1)\times k_1}, \tag{A.7}$$

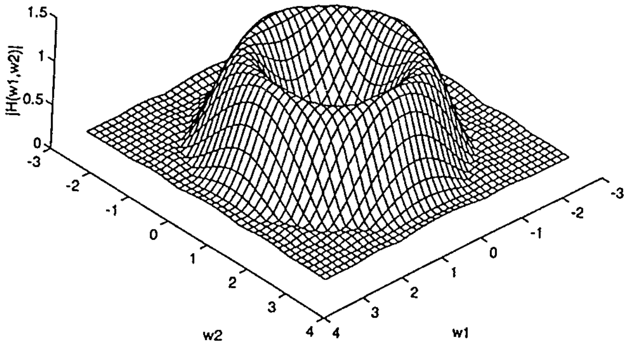
$$\mathbf{B}^T \triangleq \begin{bmatrix} \mathbf{D}_0^1 & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{D}_1^2 & \mathbf{D}_0^2 & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{D}_{m_2-1}^{m_2} & \cdots & \cdots & \mathbf{D}_0^{m_2} & \cdots & \cdots & \vdots \\ \mathbf{D}_{m_2}^{m_2+1} & \mathbf{D}_{m_2-1}^{m_2+1} & \cdots & \cdots & \mathbf{D}_0^{m_2+1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{D}_{m_2}^{k_2} & \cdots & \cdots & \cdots & \mathbf{D}_0^{k_2} \end{bmatrix} \in \mathbb{R}^{[k_1k_2-(n_1+1)(n_2+1)]\times k_1k_2}, \text{ where,} \qquad (A.8)$$

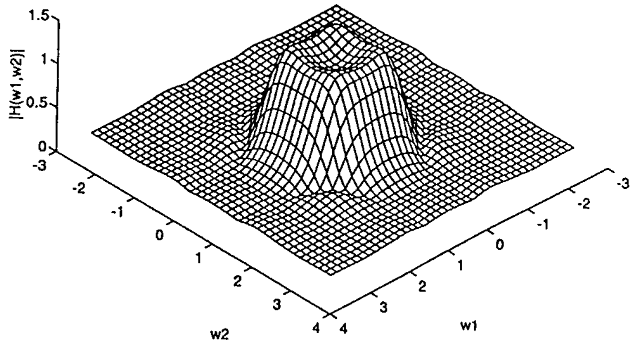
$$\mathbf{D}_i^j \triangleq \begin{bmatrix} b(n_1+1,i) & \cdots & \cdots & b(0,i) & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ b(m_1,i) & \cdots & \cdots & \cdots & b(0,i) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & b(m_1,i) & \cdots & \cdots & \cdots & b(0,i) \end{bmatrix} \in \mathbb{R}^{(k_1-n_1-1)\times k_1}, \text{ for } j \leq (n_2+1)$$

$$\triangleq \begin{bmatrix} b(0,i) & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ b(1,i) & b(0,i) & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ b(m_1-1,i) & \cdots & \cdots & b(0,i) & \cdots & \cdots & 0 \\ b(m_1,i) & b(m_1-1,i) & \cdots & \cdots & b(0,i) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & b(m_1,i) & \cdots & \cdots & \cdots & b(0,i) \end{bmatrix} \in \mathbb{R}^{k_1\times k_1}, \text{ for } j > (n_2+1)(A.9)$$

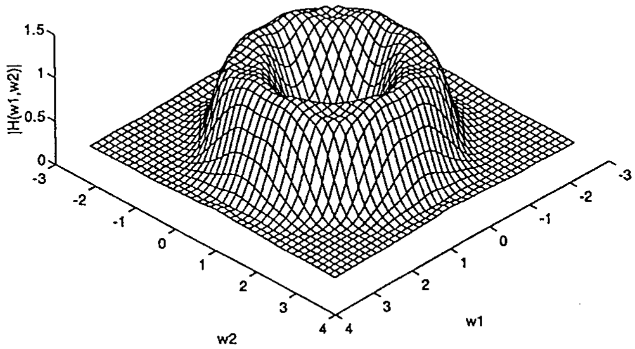
Fig. 1 : 2-D Bandpass Filter Design Example.

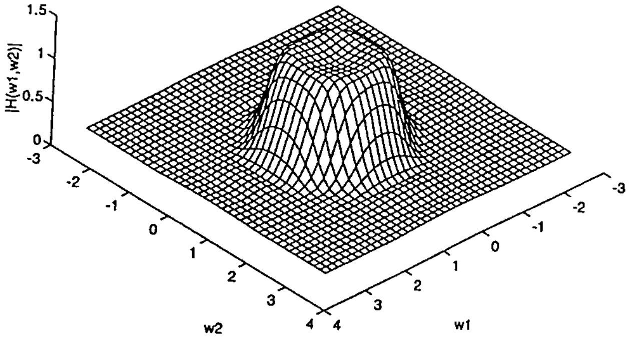
Fig. 2 : 2-D Lowpass Filter Design Example



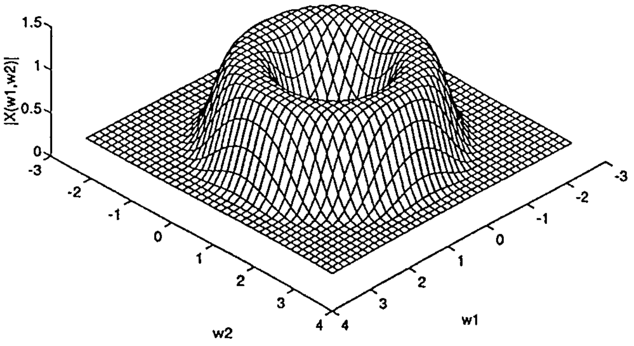


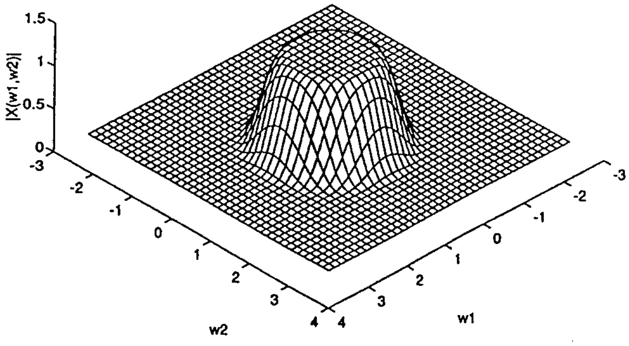
Separable-Denominator case





Optimal General Case





Prescribed Response

References

[1] J.L. Shanks, S. Treitel and J.H. Justice, "Stability and Synthesis of Two-Dimensional Recursive Filters" *IEEE Trans. Audio Electroacoust.*, Vol. AU-20, pp. 115-128, 1972.

[2] J. A. Cadzow, "Recursive Digital Filter Synthesis via Gradient Based Algorithms", *IEEE Trans. on Acoust., Speech and Signal Processing*, Vol. ASSP-24, pp. 349-355, 1976.

[3] G. A. Shaw and R. M. Mersereau, "Design, Stability and Performance of Two-Dimensional Recursive Digital Filters", *Tech. Report E21-B05-1,* Georgia Inst. of Technology School of Electrical Engg., 1979.

[4] D.E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1984.

[5] J.L.Shanks, "Recursion Filters for Digital Processing", *Geophysics*, Vol. 32, pp. 33-51, 1967.

[6] K. Steiglitz and L.E. McBride, "A Technique for Identification of Linear Systems", *IEEE Transactions on Automatic Control*, Vol. AC-10, pp. 461-464, 1965.

[7] A.G. Evans and R. Fischl, "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters", *IEEE Transactions on Audio and Electro-Acoustics*, Vol. AU-21, pp. 61-65, 1973.

[8] A.K. Shaw, "Design of Denominator Separable 2-D IIR Filters," *Signal Processing,* to be published, 1994.

[9] D.G. Luenberger, *Optimization by Vector Space Method*, John Wiley & Sons, Inc., New York, 1969.

[10] T. Hinamoto and S. Maekawa, "Spatial-Domain Design of a Class of Two-Dimensional Recursive Digital Filters", *IEEE Trans. on ASSP*, Vol.- ASSP-32, no.1, Feb.,1984.

[11] T. Hinamoto and F.W. Fairman, "Separable Denominator State Space Realization of Two Dimensional Filters using a Canonical Form", *IEEE Trans. Acoust. Speech, Sig. Proc.*, vol. ASSP-29, no. 4, pp. 846-853, 1981.

[12] T. Hinamoto and S. Maekawa, "Separable-Denominator 2-D Rational Approximation via 1-D Based Algorithm ", *IEEE Trans. Cir.Syst.*, Vol. CAS-32, pp. 989-999, Nov. 1985.

[13] T. Hinamoto, "Design of 2-D Separable-Denominator Recursive Digital Filters", *IEEE Trans. Cir.Syst.*, Vol. CAS-31, pp. 925-933, Nov. 1984.

[14] J.S.Lim, *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, NJ: Prentice Hall, 1990.

Section - 3.7 DISTRIBUTED LOOK-AHEAD : A GENERAL APPROACH FOR PIPELINING RECURSIVE DIGITAL FILTERS

SUMMARY

A new Look-Ahead (LA) scheme, *Distributed Look-Ahead* (DLA), is proposed for pipelined implementation of recursive digital filters. It is established that in case of many recursive filters, DLA can provide equivalent and *stable* implementation with reduced pipeline delay and hardware complexity, when compared with some existing LA schemes [4, 5]. The existing Scattered Look-ahead implementation [5] achieves stability at the cost of increased multiplication and latch complexities and considerable delay in output generation. The Clustered look-ahead approach can not always guarantee stability [5]. This work shows that, in order to attain stability, the output samples need not be clustered or equally scattered. Indeed, in many filter design problems, stability can be maintained by using *unequally distributed* past output samples. When compared with the scattered approach, the proposed scheme uses fewer number of pole-zero cancelations and the introduced roots are not necessarily at the same radii as the original filter poles. Hence, the proposed DLA scheme has reduced multiplication and latch complexities, higher area-efficiency and it produces outputs with reduced delays.

1. Introduction

Look-ahead pipelining is highly effective in attaining high sampling rate and computation speed for low-cost VLSI implementation of digital IIR filters [3-5,.9, 13]. Among existing LA schemes, the Clustered (CLA) or Time-Domain (TD) approach [3, 4, 9] generates the present output using contiguous past output samples whereas the Scattered (SLA) or z-domain (ZD) approach [5, 12] uses equally separated past output samples. A desirable feature of SLA is that stability is guaranteed. However, this is achieved at the cost of relatively large delay in output generation as well as increased multiplication and latch complexities to implement the numerator. On the other hand, CLA may require filter order augmentation to maintain stability, which also comes at the cost of increased hardware complexity [3].

In this paper, we propose a general Look-Ahead approach that opens up a large class of new possibilities to provide stable realizations with reduced pipeline delay and hardware complexity than the existing LA schemes. Possible stability regions for the proposed scheme are also addressed. This paper argues that, in order to attain stability in pipelined form, the output samples need not be clustered [3, 4, 9, 13] or equally scattered [5, 12]. Indeed, for many filter design problems, it is shown that stability can be maintained by using *unequally distributed* past output samples. The proposed scheme is denoted as Distributed Look-Ahead (DLA) approach, where the number of denominator coefficients can be kept same as that in the original filter (as in SLA) but the numerator and denominator orders are lower than that in SLA. Hence, the proposed DLA scheme can offer reduced multiplication and latch complexities, higher area-efficiency and it can produce outputs with reduced pipeline delay than the SLA scheme. Unlike SLA but similar to CLA, stability is not guaranteed with the proposed scheme. However, simple stability conditions and regions can be derived and have been presented for various pipeline stages. It is also demonstrated that the numerator and denominator polynomials can be factorized into lower-order polynomials, which can further simplify hardware implementation and complexity. Examples are included to demonstrate the validity of the proposed approach.

The paper is organized as follows. In Section 2, two existing LA schemes are briefly summarized. In Section 3, the general Look-Ahead scheme is proposed along with some examples. In Section 4, Stability conditions are presented for pipelining of a 2nd-order recursive filters using the proposed DLA scheme and in Section 5, several examples are provided to demonstrate that stable pipelined implementations with DLA can be achieved with reduced hardware complexity.

2. Existing Look-Ahead Schemes

Consider a recursive digital filter with L-th order numerator and N-th order Denominator of the form,

$$H(z) = \frac{\sum_{i=0}^{L} \beta_i z^{-i}}{1 - \sum_{i=1}^{N} \alpha_i z^{-i}} = \frac{B(z)}{A(z)} \tag{1}$$

which is to be implemented in pipelined form. The two major existing Look-Ahead forms are briefly outlined first.

2.1. Clustered Look-Ahead (CLA) : M-stage CLA pipelining of the filter in (1) would have the following form [3, 4, 9, 13] :

$$\frac{H_C^M(z) = 1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{L+M-1} z^{-(L+M-1)}}{1 + a_M z^{-M} + a_{M+1} z^{-(M+1)} + \cdots + a_{M+N-1} z^{-(M+N-1)}} \tag{2}$$

Note that the denominator coefficients are ordered in a clustered form. The numerator (non-recursive portion) can be implemented by $N + M$ multiplications and the denominator (recursive portion) can be implemented with N multiplications . Thus, the total multiplication complexity is $(2N + M)$ and the latch complexity is linear in M. The extra delay in producing output is M.

2.2. Scattered Look-Ahead (SLA) : An equivalent M-stage pipelining of the same N-th order recursive filter can be obtained by [5, 12],

$$H_S^M(z) = \frac{1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{N(M-1)+L} z^{-(N(M-1)+L)}}{1 + a_M z^{-M} + a_{2M} z^{-2M} + \cdots + a_{NM} z^{-NM}} \tag{3}$$

Note that the non-zero denominator coefficients are equally 'scattered'. The multiplication complexity for the non-recursive portion of the pipelined implementation is $(NM + 1)$ and that of the recursive portion is N. Thus, the total multiplication complexity is $(NM + M + 1)$ and the latch complexity is square in M. The extra delay in producing output is $(NM - N)$, which may be significant because of high order of the filter in pipelined form. However, if M is a power of 2, then using a decomposition technique the total multiplication and latch complexities can be further reduced [5].

3. The Proposed Distributed Look-Ahead Pipelining (DLA)

In this new look-ahead scheme, the filter transfer function is transformed to have the form,

$$H_D^M(z) = \frac{1 + b_1 z^{-1} + \cdots + b_{M+k_L-N+L} z^{-(M+k_L-N+L)}}{1 + a_M z^{-M} + a_{M+k_1} z^{-(M+k_1)} + a_{M+k_2} z^{-(M+k_2)} + \cdots + a_{M+k_L} z^{-(M+k_L)}} \tag{4}$$

where, $k_1, k_2, \cdots$, in general, can be arbitrary integer values with $k_L = N$, in order to keep the total number of denominator (a_i) coefficients same as in the original denominator in (1). It is easy to show that the two existing look-ahead schemes defined in equations (2) and (3) are special cases of this general M-stage look-ahead representation. Specifically, in case of CLA, $k_i = i$ and for SLA, $k_i = Mi$, with $k_L = (N-1)M$. Clearly, if the denominator order of DLA, $(M + k_L)$ is less than the order for SLA (NM), the proposed LA scheme would offer considerable hardware savings over SLA.

A comparison of the multiplication complexities of the three pipelining schemes are given in Table 1. Note that if M is a power of 2, then using a decomposition technique, the total multiplication complexity of the SLA scheme can be further reduced to $L + N + N \log_2(M)$.

4. DLA Based Pipelining of Second-Order Filter Blocks

In this section, an iterative scheme is given first for determining the coefficients for pipelining second-order ($N = 2$) filter blocks. Then several examples of DLA implementation are given for different values of M.

4.1. Iterative Scheme for Obtaining Augmentation Polynomial : Consider a second order filter transfer function,

$$H(z) = \frac{B(z)}{A(z)} = \frac{B(z)}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} \tag{5}$$

For DLA-based pipelining of this second-order filter, the only choice is, $k_1 = k_L = 2$ in (4). To determine the DLA pipelined coefficients from these serial coefficients, it may be noted that $H_D^M(z)$ must equal the original $H(z)$ and hence, $H_D^M(z)$ can be obtained by multiplying an augmentation polynomial $D(z)$ in the numerator as well as the denominator of (4), *i.e*,

$$H_D^M(z) = \frac{B(z)D(z)}{A(z)D(z)} = H(z) \tag{6}$$

where, the coefficients of $D(z) = 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_M z^{-M}$ should be selected such that the denominator possesses the desired DLA form in (4). It can be shown that the coefficients of the $D(z)$ polynomial can be found recursively using the following steps,

Initialize : $d_0 = 1$, $d_1 = -\alpha_1$ and $d_M = \frac{-\alpha_2}{\alpha_1} d_{M-1}$

Iterate : for $i = 2$ to $M - 1$

$$d_i = -\alpha_1 d_{i-1} - \alpha_2 d_{i-2} \tag{7}$$

4.2. Examples : Let the complex conjugate poles of the second order filter block be located at $z = re^{\pm j\theta}$. Then the transfer function of the second order filter would have the form,

$$H(z) = \frac{1}{1 - 2r\cos\theta z^{-1} + r^2 z^{-2}} \tag{8}$$

where, the numerator has been set to unity without any loss of generality. Using $\alpha_1 = 2r\cos\theta$ and $\alpha_2 = r^2$ in the iterations of (7), a 4-stage (M=4) DLA pipelined filter can be shown to have the form,

$$H_D^4(z) = \frac{1 + 2r\cos\theta z^{-1} + r^2(2\cos 2\theta + 1)z^{-2} + 4r^3\cos\theta\cos 2\theta z^{-3} + 2r^4\cos 2\theta z^{-4}}{1 - r^4(2\cos 4\theta + 1)z^{-4} + 2r^6\cos 2\theta z^{-6}} \tag{9}$$

Interestingly, this transfer function can be further factorized into a more convenient decomposed form as,

$$H_D^4(z) = \frac{(1 + 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r^2\cos 2\theta z^{-2})}{(1 - 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r^2\cos 2\theta z^{-2})}. \tag{10}$$

Implementation of this decomposed form allows hardware savings over its SLA counterpart. Using similar steps, it can be further shown that 3, 6 and 8-stage DLA implementations of the second-order filter have the following decomposed forms, respectively,

$$H_D^3(z) = \frac{1 + 2r\cos\theta z^{-1} + r^2(2\cos 2\theta + 1)z^{-2} + \frac{r^3}{2\cos\theta}(2\cos 2\theta + 1)z^{-3}}{1 - \frac{r^3}{2\cos\theta}(2\cos 4\theta + 2\cos 2\theta + 5)z^{-3} + \frac{r^5}{2\cos\theta}(2\cos\theta + 1)z^{-5}} \tag{11}$$

$$H_D^6(z) = \frac{(1 + 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r^2\cos 2\theta z^{-2} + r^4(4\cos^2 2\theta - 1)z^{-4})}{(1 - 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r\cos\theta z^{-1} + r^2 z^{-2})(1 + 2r^2\cos 2\theta z^{-2} + r^4(4\cos^2 2\theta - 1)z^{-4})} \tag{12}$$

$$H_D^8(z) = \frac{(1 + 2r\cos\theta z^{-1} + r^2z^{-2})(1 + 2r^2\cos\theta z^{-2} + r^4(2\cos 4\theta + 1)z^{-4} + r^6(4\cos 4\theta\cos 2\theta + 1)z^{-6})}{(1 - 2rcos\theta z^{-1} + r^2z^{-2})(1 + 2r\cos\theta z^{-1} + r^2z^{-2})(1 + 2r^2\cos 2\theta z^{-2} + r^4(2\cos 4\theta + 1)z^{-4} + r^6(4\cos 4\theta\cos 2\theta +} \tag{13}$$

A comparison of the hardware complexities between the SLA and the DLA schemes for $M = 3, 4, 6$ and 8 after their respective decompositions is given in Table 2.

From this table it is apparent that for all stages of pipelining the DLA scheme has a definite edge over the SLA scheme as far as hardware complexity is concerned. Next, the stability conditions for the DLA scheme are established for second-order filter blocks for a several values of M.

4.3. Stability conditions : Consider the general second order filter block with complex conjugate poles at $z = re^{\pm j\theta}$, represented by the transfer function in equation (8).

4.3.1. $M = 4$ **case** : The 4-stage DLA pipelined implementation of the second order filter is obtained by using the general iterative scheme discussed above. The transfer function of this is as in equation (10).

This 4-stage pipelined implementation will be stable if the roots of the factor, $(1 + 2r^2\cos 2\theta z^{-2})$ are less than unity. It can be shown that this would be true if $\theta < 0.5\cos^{-1}\left(\frac{1}{2r^2}\right)$. The region satisfying this stability condition is shown in Fig. 1 as the shaded area.

Hence, if for any 4-stage pipelined implementation of CLA produces an unstable filter, but it is found that the condition on θ stated above or in Fig. 1 is satisfied, then using the proposed DLA transformation would definitely be more appropriate than the SLA in (3), because the later would require extra hardware for implementing both the numerator and the denominator. The exact savings in hardware for this case can be found in Table 2.

4.3.2 $M = 6$ **case** : Consider the 6-stage implementation of the 2nd-order filter which has the convenient factored form shown in equation (12).

Because of the convenient quadratic factored form, the stability region is easy derive and is displayed in Fig. 2. It is interesting to note that these decompositions have simple power of 2 factors and hence, the corresponding hardware complexities are less than the SLA decomposition scheme given in [5]. The hardware savings being 2 multiplier-adder units and 8 latches (refer to Table 2).

Similar stability regions can be evaluated for other M values.

5. EXAMPLES

5.1. Example with $M = 4$: It had been shown in [5] that the second order transfer function

$$H(z) = \frac{1}{1 - 5/4z^{-1} + 3/8z^{-2}} \tag{14}$$

produces unstable filter with CLA implementation (refer to Figure 3(a)). Using the DLA formulae presented above, it can be shown that

the 4-stage implementation is given as,

$$H_D^4(z) = \frac{(1 + 1.25z^{-1} + 0.3750z^{2})(1 + 0.8125z^{-2})}{1 - 0.1595z^{-4} + 0.1143z^{-6}} \tag{15}$$

Using the SLA technique,

$$H_S^4(z) = \frac{(1 + 1.25z^{-1} + 0.3750z^{2})(1 + 0.8125z^{-2} + 0.1406z^{-4})}{1 - 0.3789z^{-4} + 0.0198z^{-8}} \tag{16}$$

Pole-zero plots of the 4-stage DLA and SLA implementations are shown in Figures 3(b) and 3(c). The DLA and the SLA implementations for this example is shown in Figures 4 and 5 respectively. They clearly demonstrate the hardware savings for the DLA case. Table-2 also shows the gain in hardware requirements with the DLA case.

5.2. Example with $M = 6$: Consider the second order transfer function

$$H(z) = \frac{1}{1 - 1.4z^{-1} + 0.5z^{-2}}. \tag{17}$$

A 6-stage CLA implementation of equation (17) yields an unstable pipelined filter. The pole-zero plot of the unstable CLA filter is shown in Figure 6(a). Using the DLA formulae given above, it is shown that the 6-stage DLA implementation is stable (refer to Figure 6(b)) and is given as,

$$H_D^6(z) = \frac{(1 + 1.4z^{-1} + 0.5z^{-2})(1 + 0.96z^{-2} + 0.6716z^{-4})}{1 - 0.4132z^{-6} + 0.1825z^{-8}} \tag{18}$$

and the corresponding SLA implementation is

$$H_S^6(z) = \frac{(1 + 1.4z^{-1} + 0.5z^{-2})(1 + 0.96z^{-2} + 0.6716z^{-4} + 0.2400z^{-6} + 0.0625z^{-8})}{1 - 0.1647z^{-6} + 0.0156z^{-12}} \tag{19}$$

The pole-zero plot for the SLA implementation is shown in Figure 6(c) and the hardware savings for the DLA case over the SLA case for this example is also given in Table-2.

Hence, in all these cases, there is no need to use SLA for stable pipelined realizations and considerable hardware savings and reduced pipelining delays can be achieved if DLA is used instead.

Pipelining Methods	Multiplication Complexity	Delay in producing First output
CLA	$L+M+N-1$	M
SLA	$NM+L$	NM
DLA	$M+k_L-N+2L+1$	$M+k_L$

Table 1: Comparison of Hardware Complexities between the Various Pipelining Techniques

Pipeline Stages	Pipeline Method	Number of Multiplier /Adder Units	Number of Latches	Delay in producing First Output
$M=3$	SLA	6	10	6
	DLA	5	8	5
$M=4$	SLA	6	14	8
	DLA	5	10	6
$M=6$	SLA	8	22	12
	DLA	6	14	8
$M=8$	SLA	8	30	16
	DLA	7	18	10

Table 2: Comparison of Hardware Complexities for DLA Pipelining for various M

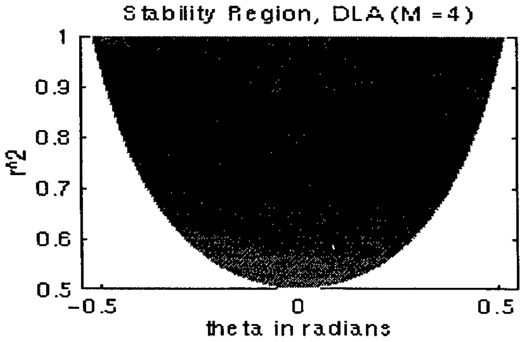


Figure 1: Stability Regions for M = 4 Case

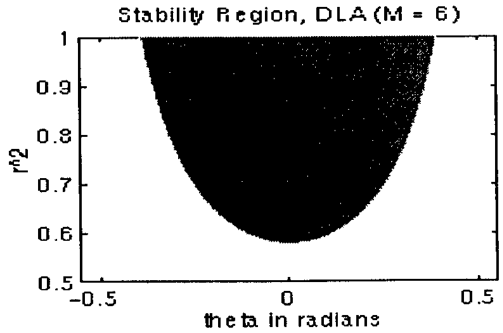


Figure 2: Stability Regions for M = 6 Case

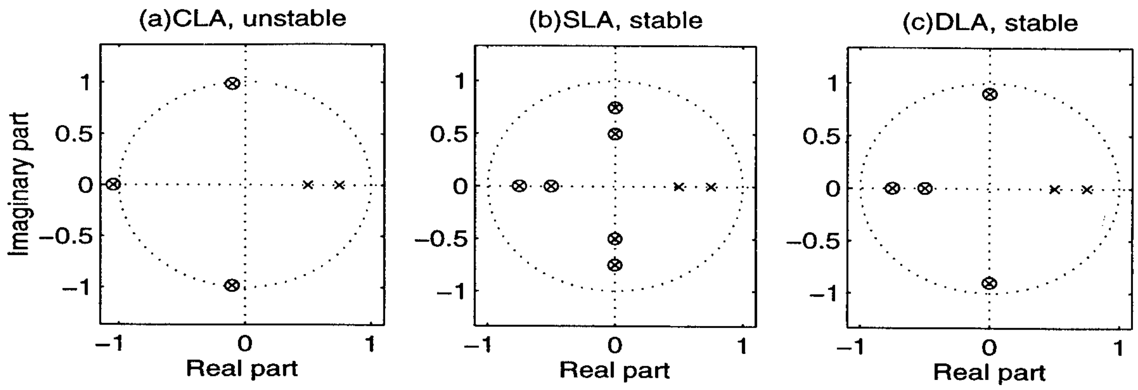


Figure 3: Pole-Zero plots for M = 4 case

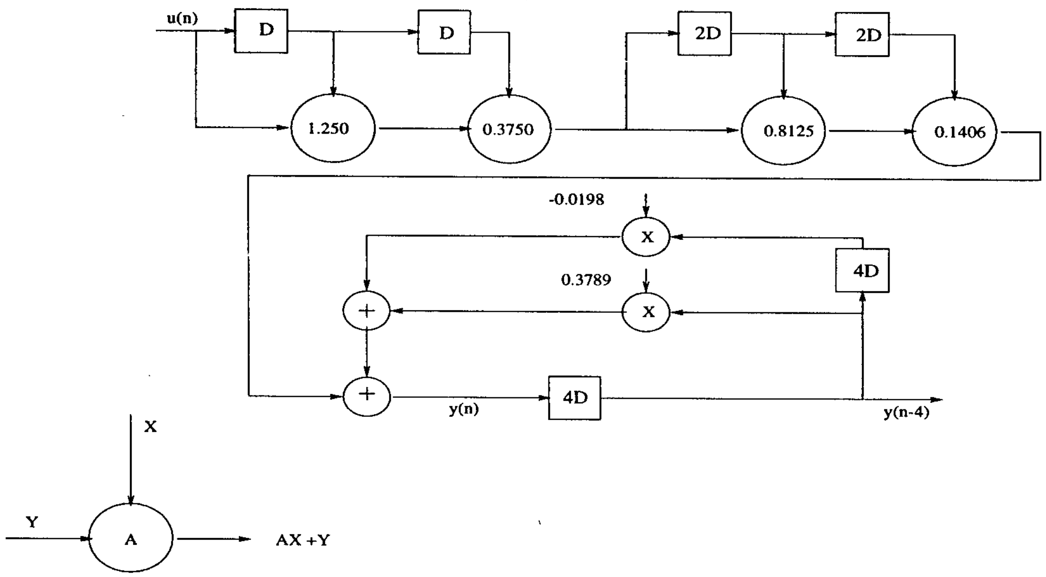


Figure 4: Implementation of a 4-stage SLA Pipelined recursive Filter Using Decomposition Technique

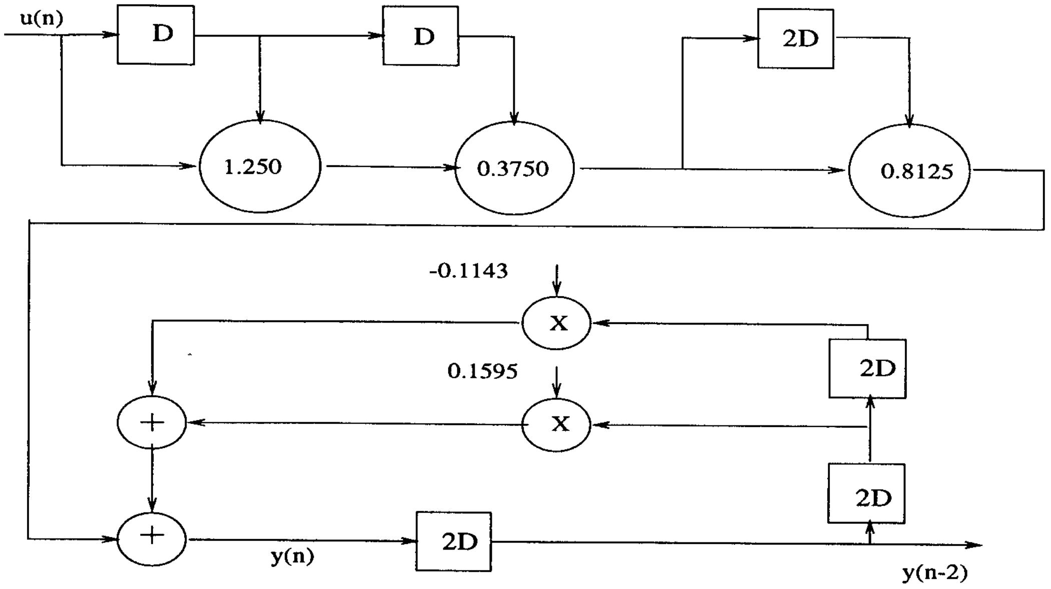


Figure 5: Implementation of a 4-stage DLA Pipelined recursive Filter Using Decomposition Technique

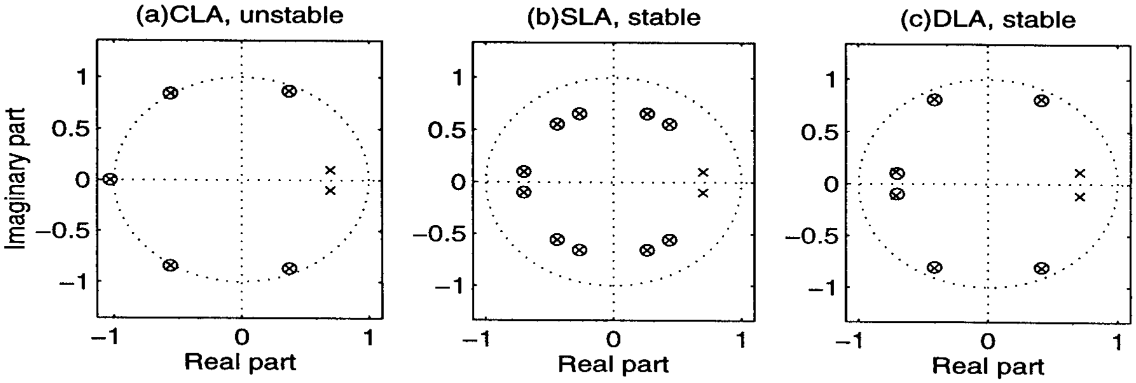


Figure 6: Pole-Zero plots for M = 6 case

References

[1] J. G. Chung and K. K. Parhi, "Design of Pipelined Lattice IIR Digital Filters," in *Proc. 25th Asilomar Conf. Signals, Syst., and Computers*, Nov. 1991.

[2] Chien-Piao Lan and Chien-Wei Jen, "Efficient Time Domain Synthesis of Pipelined recursive Filters," *IEEE Trans. Circuits Syst.,*, Vol. 41, No. 9, pp. 618-622, Sept. 1994.

[3] Y. C. Lim and B. Liu, "Pipelined Recursive Filter with Minimum Order Augmentation", *IEEE Transactions on Signal Processing*, vol.40, no. 7, pp. 1643-1651, July 1992.

[4] H.H. Loomis and B. Sinha, "High-Speed Recursive Digital Filter Realization", *Circuits, Systems and Signal Processing*, vol.3, pp. 267-294, Sept., 1984.

[5] K.K. Parhi and D.G. Messerschmitt, "Pipelining Interleaving and Parallelism in Recursive digital filters - Part I : Pipelining using Scattered Look-Ahead and Decomposition," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, pp. 1099-1117, July 1989.

[6] K.K. Parhi and D.G. Messerschmitt, "Pipelining Interleaving and Parallelism in Recursive digital filters - Part II : Pipelining Incremental Block Filtering", *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, pp. 1118-1134, July 1989.

[7] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. 42, No. 1, pp. 113-120, Jan. 1994.

[8] A. K. Shaw and M. Imtiaz, "New Look-Ahead Algorithm for Pipelined Implementation of Recursive Digital Filters," *Proceeding of ICASSP '96*, Atlanta, Georgia, May, 1996.

[9] P.M. Kogge, *The architecture of Pipelined Computers*, New York, Hemisphere Publishing Corporation, 1981.

[10] K. K. Parhi,"Algorithm Transformation Techniques for concurrent processors," Proc. IEEE, vol. 77, pp. 1879-1895, Dec. 1989.

[11] K.K. Parhi , C.Y. Wang and A.P. Brown, "Synthesis of Control Circuits in Folded pipelined DSP architectures", IEEE J. of Solid-State Circuits, vol. 27, no.1, pp. 29-43, Jan. 1992.

[12] M. A. Soderstrand, K. Chopper and B. Sinha, "Comparison of three new techniques for pipelining IIR digital filters," *Twenty-Third ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 439-443, Nov., 1984.

[13] H. B. Voelcker and E. E. Hartquist, "Digital Filtering via Block Recursion", *IEEE Trans. Audio Electroacoust.*, Vol.AU-18, pp.169-176, June, 1970.

Section - 3.8 OPTIMAL LEAST-SQUARES DESIGN OF PIPELINED RECURSIVE FILTERS IN THE TIME-DOMAIN

SUMMARY

Currently, look-ahead (LA) pipelined recursive filters are obtained primarily via transformation of a *given* un-pipelined transfer function [3-5, 9, 13]. For these approaches, it is assumed that the un-pipelined transfer function has already been designed as an intermediate step. In this Section, we present a new algorithm (OM-LA) for *direct* and *optimal* estimation of the coefficients of recursive filters in look-ahead pipelined form. OM-LA is developed by appropriate modification of a recently proposed optimal method (OM) for designing un-pipelined filters [7]. It is demonstrated that the proposed one-step approximation can achieve superior match with reduced pipelined filter order because it does not rely on pole-zero cancelations as in current LA pipelining approaches. It is also shown that the denominator polynomial can be constrained to possess any of the possible look-ahead configurations. Unlike some existing methods [1-3], OM-LA minimizes the *true* time-domain fitting error-norm between the prescribed and the estimated impulse response and produces superior results. Several examples are provided to illustrate the effectiveness of the proposed approximation algorithm.

1. Introduction Look-ahead pipelining is highly effective in attaining high sampling rate and computation speed for low-cost VLSI implementation of digital IIR filters [3-5, 9, 13]. It may be noted that the original LA schemes for pipelining recursive filters [3-5] (including DLA proposed in the previous Section), consist of two steps. First, an un-pipelined (or 'Serial') filter is assumed to be available, *i.e.*, the transfer function of the filter in serial form is assumed to have been approximated by matching some prescribed or desired specification. The LA transformations are then introduced as a second step when the filter coefficients in pipelined form are obtained by applying either CLA, SLA or DLA transformation on the serial filter coefficients. The LA schemes differ in the way order augmentation of numerator and denominator polynomials is achieved. Mathematically, the inherent transfer function remains exactly identical before and after any LA transformation is applied. The higher orders in the pipelined cases are accounted for by pole-zero cancelation which has no effect on the filter's response or its transfer function. In this part of the project, a direct approach is proposed for approximating Recursive filters having desired Look-Ahead pipelined forms.

A significant drawback of the current two-step approach to pipeline recursive filters is that the *degrees of freedom* offered by the higher orders in the pipelined filters are not exploited in any way. Moreover, for finite precision implementation using limited number of bits, pole-zero cancelation may cause numerical implementation problems. Hence, a key motivation for the later part of the paper is to explore if the look-ahead recursive filters are designed *directly* in a single step, superior approximation can be achieved at lower pipelined filter order while avoiding the pole-zero cancelation problems associated with the current two-step design process. In these regards, it may noted that frequency domain approaches have been considered in [1] while a time-domain approach had been taken in [2], though only the *modified* least-squares error criterion has been minimized. In this paper, a general theoretical framework for direct and optimal Least-Squares estimation of coefficients of pipelined digital IIR filters in the time domain is presented. The proposed approximation approach is developed by appropriate modification of a recent work by the first author on optimal time-domain approximation of recursive digital filters [7]. The true nonlinear error criterion is theoretically decoupled into two separate sub-problems of lower computational complexities. Estimation of the numerator is a linear single-step problem whereas the non-linear denominator criterion possesses a weighted quadratic form that is convenient for iterative optimization. It is shown with several examples that the proposed approach can produce pipelined filters with better match to prescribed specs with much lower filter orders and without any pole-zero cancelations.

The paper is organized as follows. In Section 2, existing LA schemes are briefly summarized. In Section 3, the one-step Look-Ahead pipelined filter approximation method is presented and in Section 4, some simulation examples are provided to illustrate the effectiveness of the direct approximation approach.

2. Transfer Functions for various Look-Ahead Schemes

Consider a recursive digital filter with L-th order numerator and N-th order Denominator of the form,

$$H(z)=\frac{\sum_{i=0}^{L}\beta_i z^{-i}}{1-\sum_{i=1}^{N}\alpha_i z^{-i}}=\frac{B(z)}{A(z)} \tag{1}$$

which is to be implemented in pipelined form. The major existing Look-Ahead forms are briefly outlined first.

2.1. Clustered Look-Ahead (CLA) M-stage CLA pipelining of this filter would have the following form: [3, 4, 9, 13],

$$H_C^M(z)=\frac{1+b_1z^{-1}+b_2z^{-2}+\cdots+b_{L+M-1}z^{-(L+M-1)}}{1+a_Mz^{-M}+a_{M+1}z^{-(M+1)}+\cdots+a_{M+N-1}z^{-(M+N-1)}} \tag{2}$$

2.2. Scattered Look-Ahead (SLA) : An equivalent M-stage pipelining of the same N-th order recursive filter can be obtained by [5, 12],

$$H_S^M(z)=\frac{1+b_1z^{-1}+b_2z^{-2}+\cdots+b_{N(M-1)+L}z^{-(N(M-1)+L)}}{1+a_Mz^{-M}+a_{2M}z^{-2M}+\cdots+a_{NM}z^{-NM}} \tag{3}$$

Note that the non-zero denominator coefficients are equally 'scattered'.

2.3. Distributed Look-Ahead Pipelining (DLA) [8, see the previous Section also]: In this new look-ahead scheme, the filter transfer function is transformed to have the form,

$$H_D^M(z)=\frac{1+b_1z^{-1}+\cdots+b_{M+k_L-N+L}z^{-(M+k_L-N+L)}}{1+a_Mz^{-M}+a_{M+k_1}z^{-(M+k_1)}+a_{M+k_2}z^{-(M+k_2)}+\cdots+a_{M+k_L}z^{-(M+k_L)}} \tag{4}$$

where, $k_1, k_2, \cdots$, in general, can be arbitrary integer values with $k_L = N$, in order to keep the total number of denominator (a_i) coefficients same as in the original denominator in (1). It is easy to show that the two existing look-ahead schemes defined in equations (2) and (3) are special cases of this general M-stage look-ahead representation. Specifically, in case of CLA, $k_i = i$ and for SLA, $k_i = Mi$, with $k_L = (N-1)M$.

2.4. General Distributed Look-Ahead Representation of Recursive Filters : In general, any of the above M-stage Look-Ahead Pipelined transfer functions can be obtained from:

$$\begin{aligned} H^M(z) &= \frac{1+b_1z^{-1}+\cdots+b_Qz^{-Q}}{1+a_Mz^{-M}+a_{M+k_1}z^{-(M+k_1)}+\cdots+a_{M+k_L}z^{-(M+k_L)}} && (5)\\ &\triangleq \frac{B^M(z)}{A(z)} && (6a)\\ &= h(0)+h(1)z^{-1}+\cdots+h(K-1)z^{-(K-1)}+\cdots && (6b) \end{aligned}$$

where, appropriate choice of Q and a set of $k_1, k_2, \cdots, k_L$ would lead to any of the desired Look-Ahead forms in (2)-(4). Note that the DLA representations in (4) and (5) differ only in the choice of the numerator order Q which need not be restricted for the approximation algorithm. In fact, by choosing Q lower than those required by (2)-(4), the total number of coefficients for the Look-Ahead representation can be reduced, if desired.

3. Proposed Method for Optimal Estimation of Coefficients of Look-Ahead Pipelined Recursive Filters

The CLA, SLA and DLA approaches to pipelining are two-step design processes which enforces "exact equality" to an already existing $H(z)$. Hence, even though the filter orders are significantly higher after any of the Look-Ahead transformations are applied, the characteristics of the filters do not change. Clearly, the original "lower order" $H(z)$ must have been obtained via some kind of approximation approach to match certain desired time-domain or frequency-domain specifications. It is well known that a superior fit can be achieved with higher filter orders. However, in case of the existing LA schemes no attempt is made to exploit the extra degrees of freedom of the higher order Look-Ahead realizations in order to achieve superior match than the original $H(z)$. In this Section, we propose to use an optimal least squares approach [7] to design the look-ahead recursive filters *directly* in a single step. The goal is to achieve superior match to the original specs with lower filter orders than otherwise would be needed with the existing two-step procedures.

Let,

$$\mathbf{h}_d = [h_d(0) \quad h_d(1) \quad \cdots \quad h_d(N-1)]^T \tag{7}$$

denote the desired impulse response (IR) of the pipelined (or un-pipelined) filter. Our goal is to estimate the a_i and b_i coefficients in (2), (3) or (4) so as to match this desired IR specification. Since the general DLA expression in equation (5) includes all the possible LA representations, we will outline only steps to determine the coefficients of the general M-stage DLA representation in (5).

Stacking the first N significant IR samples of $H^M(z)$ in (6), define,

$$\mathbf{h} = [h(0) \quad h(1) \quad \cdots \quad h(N-1)]^T . \tag{8}$$

The problem of estimating the LA coefficients to match a given $\mathbf{h}_d$ can be stated as follows,

$$min_{\mathbf{a},\mathbf{b}} \, ||\mathbf{e}||^2 \triangleq \min_{\mathbf{a},\mathbf{b}} \sum_{i=0}^{N-1} \left[h_d(i) - \frac{B(z)}{A(z)}\{\delta(i)\} \right]^2 \tag{9}$$

$$\text{where,} \delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0, \end{cases} \tag{10}$$

$$\mathbf{e} \triangleq \mathbf{h}_d - \mathbf{h}, \tag{10a}$$

$$\mathbf{a} \triangleq [1 \ a_M \ \cdots \ a_{M+k_L}]^T, \text{and} \tag{10b}$$

$$\mathbf{b} \triangleq [1 \ b_1 \ \cdots \ b_Q]^T. \tag{10c}$$

Rewriting (6) as

$$B^M(z) = H^M(z)A^M(z)$$

and equating the coefficients of equal powers of z^{-1} on both sides of this equation,

$$\begin{bmatrix} \mathbf{b} \\ \cdots \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \cdots \\ \mathbf{H}_2 \end{bmatrix} \mathbf{J}\mathbf{a}, \qquad \text{where,} \tag{11}$$

$$\mathbf{H}_1 \triangleq \begin{bmatrix} h(0) & \cdots & 0 & \cdots & 0 \\ h(1) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h(Q) & \cdots & h(0) & \cdots & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{H}_2 \triangleq \begin{bmatrix} h(Q+1) & \cdots & h(0) & \cdots & 0 \\ h(Q+2) & \cdots & h(1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h(N) & \cdots & \cdots & \cdots & h(N-M-k_L) \end{bmatrix} \tag{12}$$

The matrix $\mathbf{J}$ in (11) is necessary to constrain some of the coefficients in the denominator to be zero, and it is formed as follows: Starting with the identity matrix of size $(M + k_L + 1)$, remove all columns corresponding to the indices of those coefficients of the denominator which are zero. The remaining matrix, of size $(M+k_L+1)\times(L+2)$, becomes $\mathbf{J}$. From the bottom partition of (11), *i.e.*,

$$\mathbf{H}_2\mathbf{J}\mathbf{a} = \mathbf{0},$$

it can be shown [7] that the minimization problem stated in (9) is *theoretically equivalent* to first solving for the denominator in

$$\min_{\mathbf{a}} \quad \mathbf{a}^T\mathbf{J}^T\mathbf{H}_2^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{H}_2\mathbf{J}\mathbf{a} \tag{13a}$$

and then estimating the numerator from the top portion of (11), *i.e.*,

$$\mathbf{b} = \mathbf{H}_1\mathbf{J}\mathbf{a}. \tag{13b}$$

Note that the matrix $\mathbf{A}$ is a banded Toeplitz convolution matrix and can be defined similar to $\mathbf{H}_2$, as given in [7]. The design problem has thus been *decoupled* into two separate sub-problems of *reduced* computational complexity.

Note that the denominator criterion in (13a) has a weighted-quadratic structure where the weight matrix in the middle itself depends on the unknown coefficients. For estimating $\mathbf{a}$, an iterative algorithm has been presented in [7], where the estimates at the previous iteration is used to form the middle matrix $(\mathbf{A}^T\mathbf{A})^{-1}$. An appropriate modification of that algorithm can be used here to minimize (13a) to obtain the *optimal* estimates of the denominator $\mathbf{a}$, from which the numerator $\mathbf{b}$ can be computed using (13b).

4. Simulations on One-Step Approximation

4.2. Simulation 1

In this case, a lowpass example has been considered. In all figures, the solid line denotes the desired response and the dashed line corresponds to the response using the proposed approximation approach. Figure 1(a) shows the response of the un-pipelined filter approximation with numerator and denominator orders = 3. Note this response would remain identical if the SLA filter is obtained from it using Parhi's approach [5]. The error is $-9.9dB$. Figure 1(b) shows the response of the CLA filter approximation with $Q = 6, M = 4k_1 = 1$ and $k_2 = 2$ and has an error of $-19.5dB$. Figure 1(d) shows the response of the DLA filter with $Q = 6, M = 4, k_1 = 2$ and $k_2 = 4$ with an error of $-32.5dB$. Finally, Figure 1(c) shows the response of the SLA filter designed *directly* by the OM-LA. For Figure 1(c) $-80.7dB$. It is evident from the error values and a comparison of Figures 1(a) and 1(c) that the SLA filter designed directly by OM-LA is much superior to that of Parhi [5]. The respective pole-zero plots of the filters are shown alongside. Note that the OM-LA does not produce cancelling poles and zeros.

4.2. Simulation 2

A notch filter example has been considered for this example. Figure 2(a) shows the response of the un-pipelined filter with numerator order, $L = 8$ and denominator order, $N = 10$. The responses with the CLA , SLA and DLA approximations are shown in Figures 2(b), (c) and (d) respectively. Note the filter order and hardware requirements for the SLA filter would be extremely high even for the pipelining stage of $M = 3$. Table 1 shows a comparison of the hardware used in the three cases. From Table 1 and the responses in Figure 2, it is apparent that stable CLA or DLA approximations can be achieved with excellent match and much reduced hardware requirements than SLA.

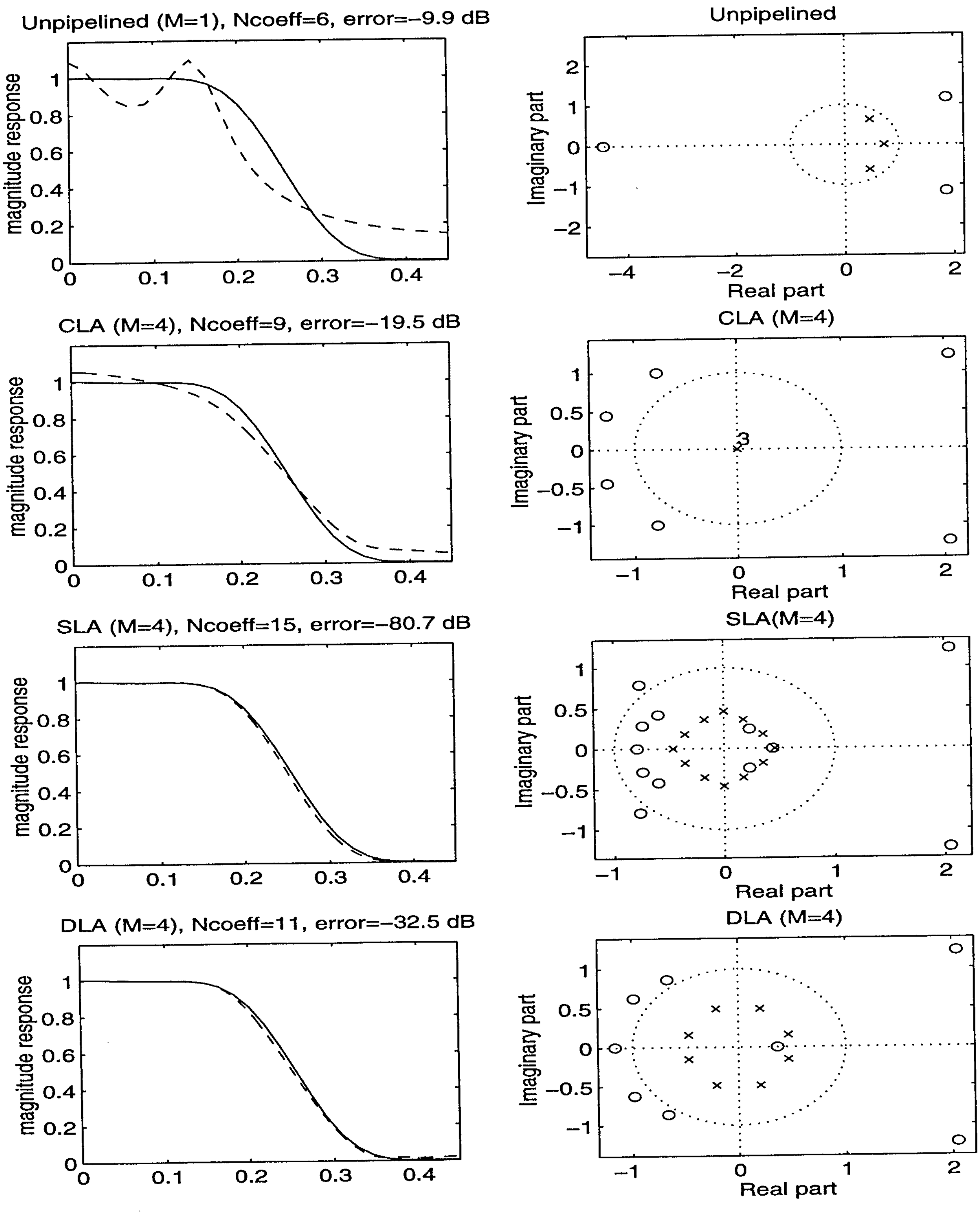


Figure 1: Lowpass Filter: (a) Response of an un-pipelined filter (b) Response of the CLA filter (c) Response of the SLA filter obtained directly through OM-LA (d) Response of the DLA filter, Legend:- - desired, – OM-LA

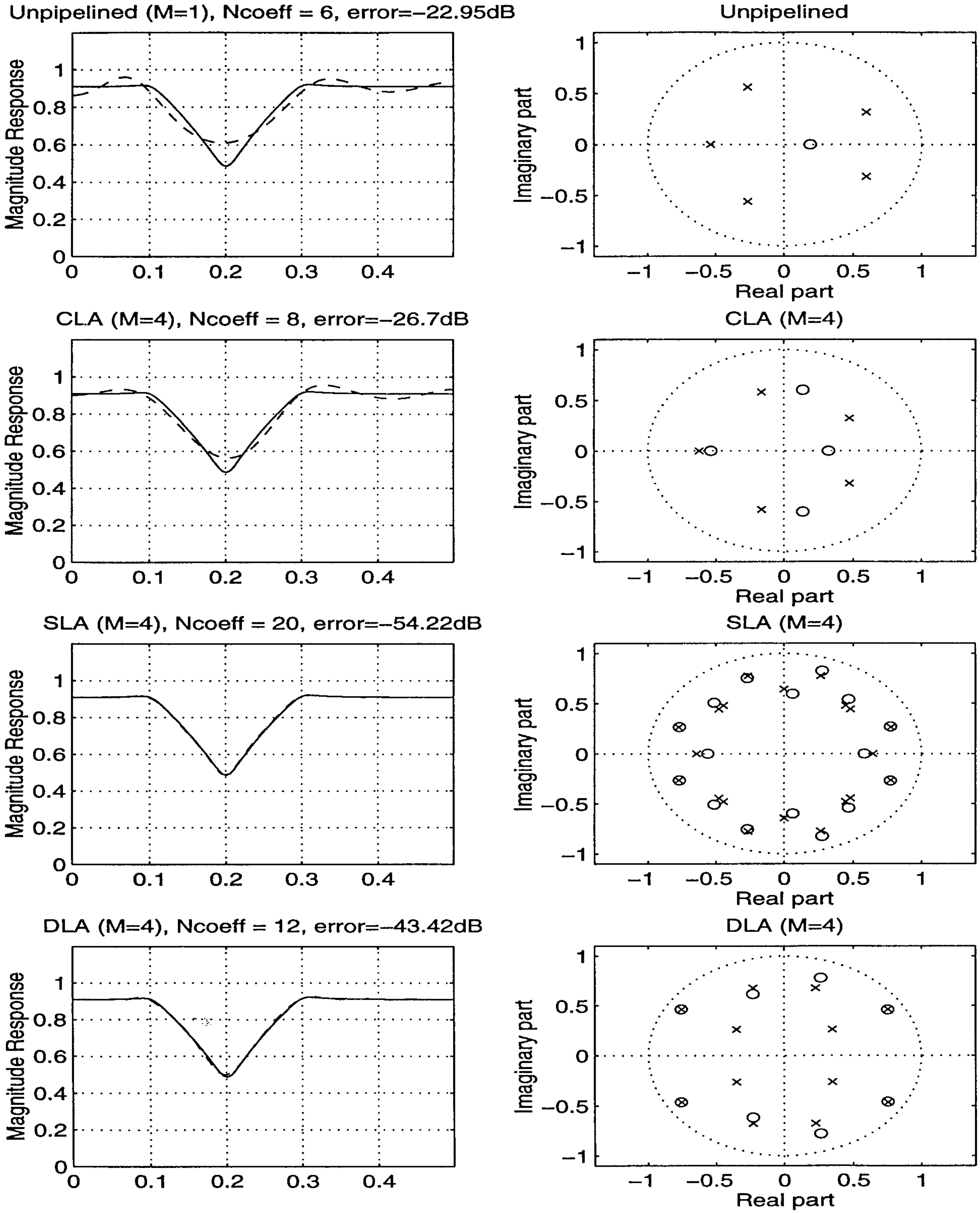


Figure 2: Notch Filter: (a) Response of an un-pipelined filter (b) Response of the CLA filter (c) Response of the SLA filter obtained directly through OM-LA (d) Response of the DLA filter, Legend:- - desired, – OM-LA

References

[1] J. G. Chung and K. K. Parhi, "Design of Pipelined Lattice IIR Digital Filters," in *Proc. 25th Asilomar Conf. Signals, Syst., and Computers*, Nov. 1991.

[2] Chien-Piao Lan and Chien-Wei Jen, "Efficient Time Domain Synthesis of Pipelined recursive Filters," *IEEE Trans. Circuits Syst.,*, Vol. 41, No. 9, pp. 618-622, Sept. 1994.

[3] Y. C. Lim and B. Liu, "Pipelined Recursive Filter with Minimum Order Augmentation", *IEEE Transactions on Signal Processing*, vol.40, no. 7, pp. 1643-1651, July 1992.

[4] H.H. Loomis and B. Sinha, "High-Speed Recursive Digital Filter Realization", *Circuits, Systems and Signal Processing*, vol.3, pp. 267-294, Sept., 1984.

[5] K.K. Parhi and D.G. Messerschmitt, "Pipelining Interleaving and Parallelism in Recursive digital filters - Part I : Pipelining using Scattered Look-Ahead and Decomposition," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, pp. 1099-1117, July 1989.

[6] K.K. Parhi and D.G. Messerschmitt, "Pipelining Interleaving and Parallelism in Recursive digital filters - Part II : Pipelining Incremental Block Filtering", *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 37, pp. 1118-1134, July 1989.

[7] A. K. Shaw, " Optimal Identification of Discrete-Time Systems from Impulse Response Data," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. 42, No. 1, pp. 113-120, Jan. 1994.

[8] A. K. Shaw and M. Imtiaz, "New Look-Ahead Algorithm for Pipelined Implementation of Recursive Digital Filters," *Proceeding of ICASSP '96*, Atlanta, Georgia, May, 1996.

[9] P.M. Kogge, *The architecture of Pipelined Computers*, New York, Hemisphere Publishing Corporation, 1981.

[10] K. K. Parhi,"Algorithm Transformation Techniques for concurrent processors," Proc. IEEE, vol. 77, pp. 1879-1895, Dec. 1989.

[11] K.K. Parhi , C.Y. Wang and A.P. Brown, "Synthesis of Control Circuits in Folded pipelined DSP architectures", IEEE J. of Solid-State Circuits, vol. 27, no.1, pp. 29-43, Jan. 1992.

[12] M. A. Soderstrand, K. Chopper and B. Sinha, "Comparison of three new techniques for pipelining IIR digital filters," *Twenty-Third ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 439-443, Nov., 1984.

[13] H. B. Voelcker and E. E. Hartquist, "Digital Filtering via Block Recursion", *IEEE Trans. Audio Electroacoust.*, Vol.AU-18, pp.169-176, June, 1970.

Section - 3.9 PIPELINED LOOK-AHEAD IMPLEMENTATION OF A CLASS OF 2-D IIR FILTERS

SUMMARY

In the previous section we have presented a new scheme (referred to as *distributed look-ahead*) which is a compromise between the two existing look-ahead approaches for high speed implementation of 1-D Recursive Digital filters. To date neither the Scattered Look-ahead nor the Distributed scheme has so far been utilized for 2-D IIR filter implementation, primarily because the 1-D stability properties of these LA schemes do not necessarily translate to general 2-D IIR filters. The primary focus of this paper is to demonstrate that for a special but very important class of 2-D IIR filters, namely for Denominator Separable configurations, the benefits of these stable look-ahead schemes can indeed be taken advantage of. The efficiency of the proposed implementations and the reductions in multiplication and delays are demonstrated with some examples.

I. Introduction :

Two-dimensional (2-D) IIR filters have many practical applications, such as in radar, digital image processing, remote sensing, etc. Processing time and throughput delay are two of the major problems for implementing 2-D digital IIR filters. Look-Ahead (LA) pipelining has been found to be highly effective for attaining high sampling rate and high computation speed for low-cost VLSI implementation of recursive digital filters [1-6]. In particular, the Clustered Look-Ahead (CLA) scheme has been utilized for implementing both 1-D and 2-D IIR filters [2]. However, it is known that even for the 1-D case CLA can not assure stability [2]. In order to avoid the stability problems of CLA, several other LA schemes have been proposed, namely, Scattered Look-Ahead (SLA) [2], Minimum Augmentation CLA (MACLA) [3] and Distributed Look-Ahead (DLA) [9]. To the best of our knowledge these later schemes have not so far been utilized for 2-D IIR filter implementation, primarily because the 1-D stability properties of these LA schemes do no necessarily translate to general 2-D IIR filters. The primary focus of this paper is to demonstrate that for a special but very important class of 2-D IIR filters, namely for Denominator Separable configurations, the benefits of these stable look-ahead schemes can indeed be taken advantage of.

Separable-Denominator 2-D IIR filters have considerable practical applications. Firstly, many commonly used 2-D filters such as, Gaussian, Laplacian-Gaussian, Lowpass, Bandpass, are known to possess symmetric spatial response and hence, many of these filters inherently conform to denominator-separable transfer functions [7, 8, 11-13]. Second, a general 2-D filter can be approximated by a denominator-separable filter [12, 13]. Thirdly, the design of 2-D separable-denominator filters are much easier and each of the 1-D denominators can be implemented using highly modular structures [11]. But most importantly, the stability tests for these filters are simpler and identical to those for 1-D filters.

Direct form realizations of 2-D denominator separable IIR filters have been attempted [11, 13], but have certain speed disadvantages. Block filtering techniques [9] with a combination of scattered look-ahead and decomposition based pipelining [2] can be used as an approach for implementation of 2-D denominator separable filters; this approach being carried out on each of the two separated domains of the denominator of the 2-D denominator separable filter. However the state update in the Block filters is based on clustered look-ahead [9] approach which does not necessarily guarantee stability and at the same time block structures are highly complex.

In this paper, we show that, utilizing the SLA and DLA pipelining techniques, high-speed modular implementation of separable-denominator stable 2-D IIR filters is indeed feasible. It may be noted that the various stable LA schemes recast the way the the output is generated by appropriate placement of the 1-D poles. The numerator does not play any role in stability considerations. Hence, if the original 1-D factors are stable to begin with, application of any of the stable LA schemes to individual 1-D factors would also maintain stability for the overall 2-D filter. Block Processing for further speed-up is also feasible for the proposed architectures.

II. Look-Ahead pipelining for 2-D Denominator-Separable IIR filters

A denominator separable 2-D IIR filter transfer function is given by

$$H(z_1,z_2) = \frac{}{D(z_1,z_2)} = \frac{A(z_1,z_2)}{B(z_1)C(z_2)} = \frac{\sum_{i=0}^{M_1}\sum_{j=0}^{M_2} a_{i,j} z_1^{-i} z_2 - j}{\sum_{i=0}^{N_1} b_i z_1^{-i} \sum_{j=0}^{N_2} c_j z_2^{-j}} \tag{1}$$

II.1. Clustered Look-Ahead Pipelining [1, 2, 6]:

M-stage pipelining of an (N_1, N_2)-th order separable-denominator 2-D IIR filter can be represented as,

$$\begin{aligned} H(z_1,z_2) &= \frac{\sum_{i=0}^{M_1}\sum_{j=0}^{M_2} a_{i,j} z_1^{-i} z_2 - j}{\sum_{i=0}^{N_1} b_i z_1^{-i} \sum_{j=0}^{N_2} c_j z_2 - j} & (2a) \\ &= \frac{A(z_1,z_2)(1 + d_{1,0}z_1^{-1} + d_{0,1}z_2^{-1} + d_{1,1}z_1^{-1}z_2^{-1} + \cdots + d_{M-1,M-1}z_1^{-(M-1)}z_2^{-(M-1)})}{(1 + b_M z_1^{-M} + b_{M+1}z_1^{-(M+1)} + \cdots + b_{M+N_1-1}z_1^{M+N_1-1})(1 + c_M z_2^{-M} + c_{M+1}z_2^{-(M+1)} + \cdots + c_{M+N_2-1}z_2^{M+N_2-1})} & (2b) \end{aligned}$$

where, $d_{i,j}$'s denote the coefficients of the extra numerator polynomial introduced due to the pipelining of the denominator [2, 6]. The multiplication complexity is $(2N_1+M)(2N_2+M)$ and the latch complexity is linear in M. The extra delay in producing output is M for each domain. It may be noted that this scheme may suffer from the same stability problems of its 1-D counterparts [2]. A Clustered approach that guarantees stability with minimum augmentation of order (MACLA) may be utilized [3]. However, finding the coefficients for MACLA appears to be somewhat cumbersome. Hence, in our examples we will use SLA and DLA which are briefly outlined next.

II.2. Scattered Look-Ahead Pipelining [2]:

An equivalent M-stage pipelining of the same (N_1, N_2)-th order recursive filter can be obtained by,

$$H(z_1,z_2) = \frac{A(z_1,z_2)(1 + d_{1,0}z_1^{-1} + d_{0,1}z_2^{-1} + d_{1,1}z_1^{-1}z_2^{-1} + \cdots + d_{N_1(M-1),N_2(M-1)}z_1^{-N_1(M-1)}z_2^{-N_2(M-1)})}{(1 + b_M z_1^{-M} + b_{2M}z_1^{-2M} + \cdots + b_{N_1M}z_1^{-N_1M})(1 + c_M z_2^{-M} + c_{2M}z_2^{-2M} + \cdots + c_{N_2M}z_2^{-N_2M})} \tag{3}$$

The total multiplication complexity is $(N_1M+M+1)(N_2M+M+1)$ and the latch complexity is square in M in each domain. The extra delay in producing output is $N_1(M-1)+N_2(M-1)$. However, if M is a power of 2, then using a decomposition technique [2],the total multiplications can be reduced to $(2N_1+N_1\log_2 M+1)(2N_2+N_2\log_2 M+1)$.

II.3. Distributed Look-Ahead Pipelining [9] :

For this recently proposed look-ahead scheme, the filter transfer function (2a) is transformed to the form

$$H(z_1,z_2) = \frac{A(z_1,z_2)(1 + d_{1,0}z_1^{-1} + d_{1,1}z_1^{-1}z_2^{-1} + \cdots + d_{M+k_L,M+k_L}z_1^{-(M+k_L)}z_2^{-(M+k_L)})}{(1 + b_M z_1^{-M} + b_{M+k_1}z_1^{-(M+k_1)} + \cdots + b_{M+k_L}z_1^{-(M+k_L)})(1 + c_M z_2^{-M} + c_{M+k_1}z_2^{-(M+k_1)} + \cdots + c_{M+k_L}z_2^{-(M+}} \tag{4}$$

where, $k_1, k_2, \cdots$ are integers. It is easy to show that the look-ahead schemes in (2) and (3) are special cases of this general M-stage look-ahead approach. In [9], the stability conditions for a few low-M cases have been presented. The examples below will demonstrate that, when compared with SLA, this new scheme can produce stable implementations with lower multiplication and latch complexities and reduced output delay. In our examples, $k_L = N_1 = N_2 = N$ will be used. Clearly, if $(M + k_L) < NM$, this scheme would provide considerable savings over the scattered approach.

III. Examples of Look-Ahead Pipelined Implementation of 2-D Denominator-Separable IIR Filters:

In this Section, we show that for 2-D Denominator-Separable IIR filters, any of the pipelining schemes discussed above can be easily adopted for high-speed implementation. Specifically, denominator-separable designs of many commonly occurring 2-D filters such as, Gaussian, Laplacian, Lowpass and Bandpass provide excellent match [7]. We provide two examples using Gaussian and Laplacian filters to show that, with separable denominators in two domains, the recursive sections can be easily transformed to pipelinable forms using any of the forms in (2), (3) or (4).

III.1. Example 1 : Gaussian Filter Implementation

Consider a 4-stage ($M = 4$) implementation of a (4,4) order 2-D Gaussian IIR filter with the following coefficients,

$$A(z_1, z_2) = \begin{bmatrix} 0.00936980949545 & -0.00126723735355 & 0.00825275276870 & 0.00489009121304 \\ -0.0012672373559 & 0.000171389878746 & -0.0011161589361 & -0.0006613695040 \\ 0.00825275276913 & -0.00111615893404 & 0.00726887011928 & 0.00430710078143 \\ 0.00489009121014 & -0.00066136950242 & 0.00430710077865 & 0.00255213215053 \end{bmatrix} \tag{5}$$

$$\begin{aligned} B(z_1) &= 1 - 2.2195z_1^{-1} + 2.0846z_1^{-2} - 0.9754z_1^{-3} + 0.19065z_1^{-4} && (6a) \\ &= (1 - 0.96486z_1^{-1} + 0.4557z_1^{-2})(1 - 1.2546z_1^{-1} + 0.41834z_1^{-2}) && (6b) \\ &\triangleq B_1(z_1)B_2(z_1) && (6c) \end{aligned}$$

In polar form ($z_i = r_i e^{\pm j\theta_i}$), $B(z_1)$ has four roots (poles) with radii, $r_1 = 0.6758$ and $r_2 = 0.6467$ and angles, $\theta_1 = \pm 44.38$ and $\theta_2 = \pm 14.18$, respectively. Because of symmetry, $C(z_2)$ has identical coefficients as $B(z_1)$. Hence, the poles of $C(z_2)$ are identical to those of $B(z_1)$. Each second order of $B(z_1)$ and $C(z_2)$ are pipelined separately. Incidentally, the clustered approach produced unstable filter in this case (refer to the pole-zero plot in Fig. 1). Hence, the clustered implementation given in [6] will not be suitable for pipelining this particular filter. The scattered approach can certainly be used, but the coefficients in (6) also satisfies the stability conditions given in [9]. Hence, the recently proposed distributed look-ahead scheme [9] would provide stable filters with considerable hardware savings. Next we provide the coefficients for equivalent pipelined filter implementations using both SLA and DLA schemes.

Scattered Look-Ahead Implementation :

It can be shown that 4-stage pipelining of the second-order factors in (6b) have the following forms,

$$\begin{aligned} \frac{1}{B_1(z_1)} &= \frac{(1 + 0.0195z_1^{-2} + 0.2077z_1^{-4})(1 + 0.9649z_1^{-1} + 0.4557z_1^{-2})}{1 + 0.4150z_1^{-4} + 0.0431z_1^{-8}} \quad \text{and} && (7a) \\ \frac{1}{B_2(z_1)} &= \frac{(1 + 0.7374z_1^{-2} + 0.1750z_1^{-4})(1 + 1.2546z_1^{-1} + 0.4183z_1^{-2})}{1 - 0.1938z_1^{-4} + 0.0306z_1^{-8}} && (7b) \end{aligned}$$

Pipelining of $C_1(z_2)$ and $C_2(z_2)$ would also produce identical coefficients.

Distributed Look-Ahead Implementation :

In this case, 4-stage pipelining of the second-order factors in (6b) would have the following forms,

$$\frac{1}{B_1(z_1)} = \frac{(1+0.9649z_1^{-1}+0.4557z_1^{-2})(1+0.0195z_1^{-2})}{1+0.2073z_1^{-4}+0.0040z_1^{-6}} \quad \text{and} \tag{8a}$$

$$\frac{1}{B_2(z_1)} = \frac{(1+1.2546z_1^{-1}+0.4183z_1^{-2})(1+0.7374z_1^{-2})}{1-0.3688z_1^{-4}+0.1291z_1^{-6}} \tag{8b}$$

Because of symmetry, $C_1(z_2)$ and $C_2(z_2)$ will also have identical coefficients.

It may be emphasized here that both SLA and DLA implementations in (7) and (8), respectively, produce stable high speed structures. However, comparing the denominators as well as the numerator factors in (7) and (8), it easy to see that the recently proposed DLA scheme provides considerable savings in multiplication and latch complexities and reduced delay in output generation. The pole locations with 4-stage lookahead for this example are shown in Fig. 2. In Figs. 3 and 4, the signal flow diagrams are given for a pair of 2nd order blocks in two domains for 4-stage scattered and distributed look-ahead pipelining, respectively. Comparing the number of delays and the multipliers in Figs. 3 and 4 also it is obvious that DLA can offer reduced complexity than SLA.

III.2. Example 2 : Laplacian Filter Implementation

In this case, a (4,4) order 2-D Laplacian IIR filter with the following coefficients are considered,

$$A(z_1,z_2) = \left[\begin{matrix} -0.00234823839532 & -0.00391545411710 & -0.02106505604166 & -0.00234522754996 & -0.0230375942386 \\ -0.00373345042666 & 0.00476642973501 & -0.01976926214265 & 0.01122239739174 & -0.0264271437586 \\ -0.02194869729514 & -0.02059991460608 & -0.04640872036691 & 0.14204249387362 & -0.1034676446472 \\ 0.00165842866327 & 0.01338997088113 & 0.1370918579269 & 0.1639249127551 & 0.09669575185083 \\ -0.02470285791456 & -0.02800618927861 & -0.09758570689236 & 0.11045024575180 & -0.1501644414553 \end{matrix}\right. \tag{9}$$

$$\begin{aligned} B(z_1) &= 1-1.64180z_1^{-1}+1.50063z_1^{-2}-0.80133z_1^{-3}+0.21866z_1^{-4} && (10a)\\ &= (1-0.4852z_1^{-1}+0.5146z_1^{-2})(1-1.1566z_1^{-1}+0.4249z_1^{-2}) && (10b)\\ &\triangleq B_1(z_1)B_2(z_1) && (10c) \end{aligned}$$

The radii and the angles of the roots of $B(z_1)$ are, $r_1 = 0.71736$, $r_2 = 0.65185$ and $\theta_1 = \pm 70.23$ and $\theta_2 = \pm 27.48$, respectively. The poles of $C(z_2)$ also have the same values. Similar to Example 1, each second order of $B(z_1)$ and $C(z_2)$ are pipelined separately. The clustered approach of [6] again produced unstable filter in this case (due to $B_2(z_1)$). However, in this case also, the coefficients in (10b) satisfied the stability conditions of DLA [9]. Hence, both the scattered and distributed look-ahead pipelining methods can provide stable filters with DLA providing more hardware savings than SLA. The coefficients for scattered and distributed pipelining implementations for this example are given next.

Scattered Look-Ahead Implementation :

It can be shown that 4-stage pipelining of the second-order factors in (10b) have the following forms,

$$\frac{1}{B_1(z_1)} = \frac{(1-0.7938z_1^{-2}+0.2648z_1^{-4})(1+0.4852z_1^{-1}+0.5146z_1^{-2})}{1-0.1005z_1^{-4}+0.0701z_1^{-8}} \tag{11a}$$

$$\frac{1}{B_2(z_1)} = \frac{(1+0.4879z_1^{-2}+0.1805z_1^{-4})(1+1.1566z_1^{-1}+0.4249z_1^{-2})}{1+0.1231z_1^{-4}+0.0326z_1^{-8}} \tag{11b}$$

$C_1(z_2)$ and $C_2(z_2)$ would also have identical forms.

Distributed Look-Ahead Implementation :

4-stage pipelining of the second-order factors in (10b) of the Laplacian filter when pipelined by distributed look-ahead pipelining leads to the following,

$$\frac{1}{B_1(z_1)} = \frac{(1+0.4852z_1^{-1}+0.5146z_1^{-2})(1+0.7938z_1^{-2})}{1-0.3653z_1^{-4}-0.2102z_1^{-6}} \quad (12a)$$

$$\frac{1}{B_2(z_1)} = \frac{(1+1.1566z_1^{-1}+0.4249z_1^{-2})(1+0.4879z_1^{-2})}{1-0.0575z_1^{-4}+0.0881z_1^{-6}} \quad (12b)$$

$C_1(z_2)$ and $C_2(z_2)$ will also have identical forms. The pole locations for the 4-stage look-ahead schemes for this Laplacian example are given in Fig. 5 which clearly shows that both SLA and DLA provide stable implementations. Equations (11) and (12) further show that the DLA given in [9] scheme can provide reduced complexity and reduced delay for stable implementation of high-speed separable-denominator 2-D IIR filters. More example will be included in the paper incorporating block processing [10] for further improvement in throughput.

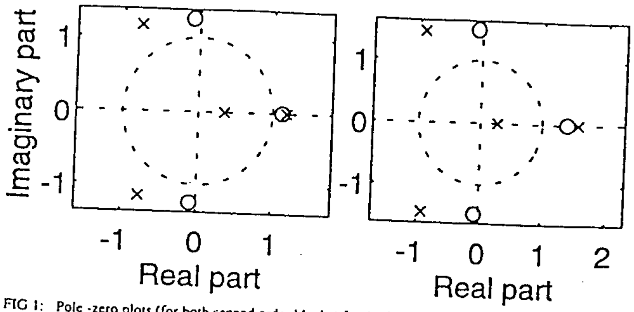


FIG 1: Pole -zero plots (for both second order blocks) for the 2-D denominator separable Gaussian filter using clustered pipelining method.

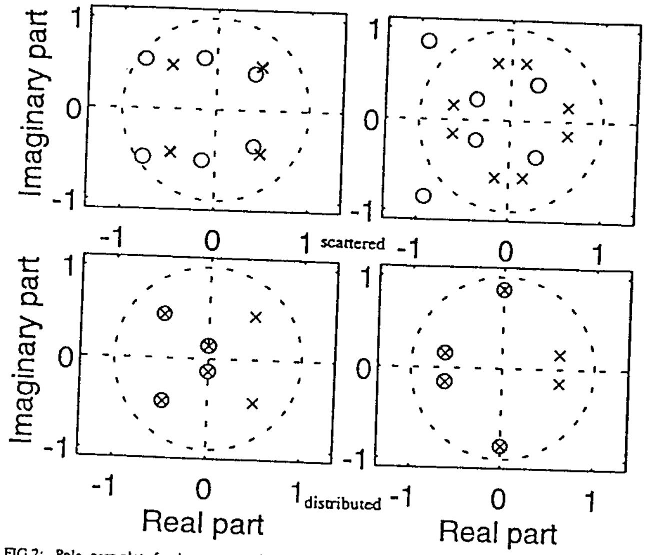


FIG 2: Pole -zero plots for the two second-order blocks using 4-stages of scattered and distributed pipelining methods for the Gaussian 2-D denominator separable filter.

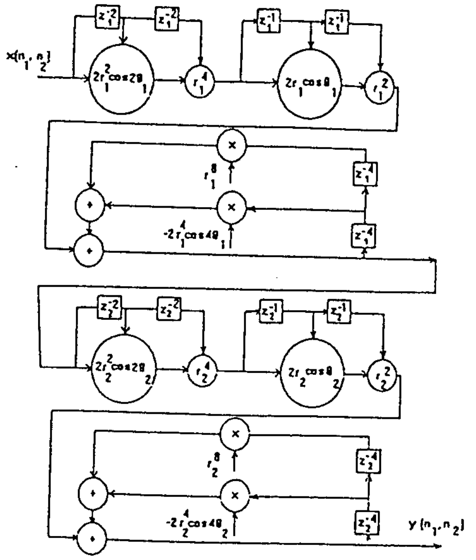


FIG 3: Implementation of the second-order blocks in both domains after 4-stages of scattered pipelining with decomposition technique inside the recursive loop.

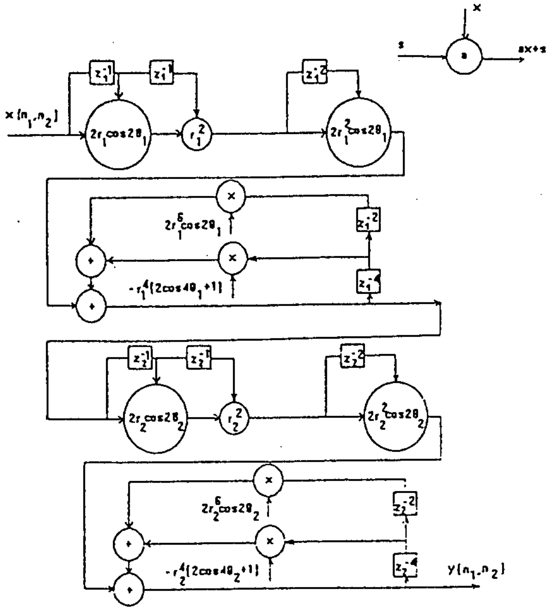


FIG 4: Implementation of the second-order blocks in both domains after 4-stages of distributed pipelining with decomposition technique inside the recursive loop.

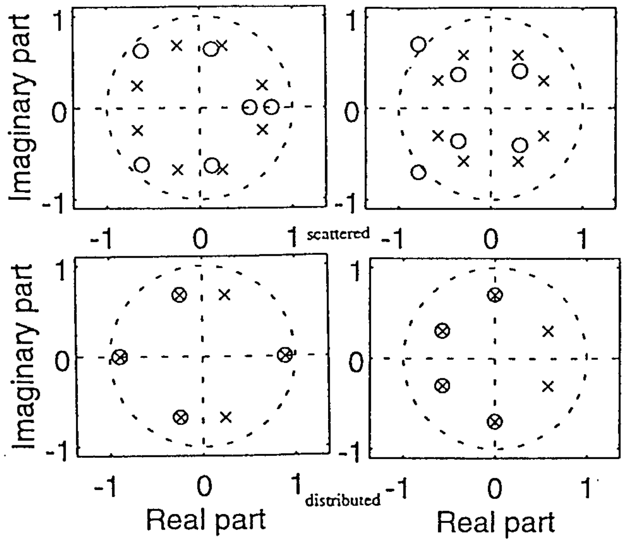


FIG 5: Pole -zero plots for the two second-order blocks using 4-stages of scattered and distributed pipelining methods for the Laplacian 2-D denominator separable filter.

References

[1] H. H. Loomis and B. Sinha, "High Speed Recursive Digital Filter Realization",*Circuits, Syst., Signal Processing*, Vol.3, no.3, pp. 267-294, 1984.

[2] K. K. Parhi and D. Messerschmitt, "Pipeline Interleaving and Parallelism in Recursive Digital Filters-Part I : Pipelining Using Scattered Look-Ahead and Decomposition," *IEEE Trans. Acoust.,Speech and Signal Processing.*, Vol. 37, pp. 1099-1117, July 1989.

[3] Y. C. Lim and B. Liu, "Pipelined Recursive Filter with Minimum Order Augmentation," *IEEE Trans. Cir. Syst.*, Vol.40, pp.1643-1651, July 1992.

[4] H. B. Voelcker and E. E. Hartquist, "Digital Filtering via Block Recursion," *IEEE Trans. Audio Electroacoust.*, Vol.AU-18, pp.169-176, June 1970.

[5] P. M. Kogge and H. S. Stone, "A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations," *IEEE Trans. Comput.*, Vol. C-22, pp. 786-793, Aug. 1973.

[6] K. K. Parhi and D. G. Messerschmitt, "Concurrent Architectures for Two-Dimensional Recursive Digital Filtering," *IEEE Trans. Cir. Syst.*, Vol. 36, pp. 813-829, June 1989.

[7] A. K. Shaw, "An Optimal Method For Identification of Pole-Zero Transfer Functions". *IEEE Trans. ASSP*, vol. 42, no. 1, pp. 113-120, Jan. 1994.

[8] D. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[9] A. K. Shaw, M. Imtiaz and S. Pokala, "New Look-Ahead Algorithms and Architectures for Pipelined Recursive Filters," Submitted to, *ICASSP-95.*

[10] K. K. Parhi and D. G. Messerschmitt, "Pipeline Interleaving and Parallelism in Recursive Digital Filters-Part II: Pipelined Incremental Block Filtering," *IEEE Trans. on Signal Processing*, vol. 37, no.7, pp. 1118-1135, July 1989.

[11] D. Raguramireddy and R. Unbehauen, "Highly Modular Systolic Structures for Denominator Separable 2-D Recursive Filters," *IEEE Trans. on Signal Processing*, vol. 39, no. 12, pp. 2725-2728, Dec. 1991.

[12] M. Yahia and W. E. Alexander, "Multiprocessor Implementation of 2-D Denominator Separable Digital Filter for Real-Time Processing," *IEEE Trans. on Signal Processing*, vol. 37, no.6, pp. 872-881, June 1989.

[13] B. G. Mertzois and A. N. Venetsanopoulos, "VLSI Implementation of Two-Dimensional Digital Filters via Two-Dimensional Filter Chips," *IEEE Journal of Solid-State circuits*, vol. SC-21, no.1, pp. 129-139, Feb. 1986.